2025 IEEE INTERNATIONAL CONFERENCE ON
ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (ICASSP 2025)

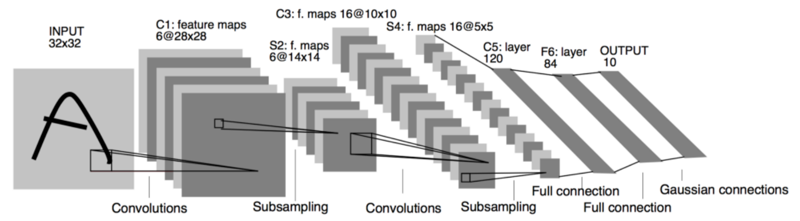
Towards Time-Frequency Deformation Bounds for Deep Convolutional Neural Networks

Albert Chua

April 10, 2025

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
  - Perform a pooling operation (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."

A Convolutional Feature Extractor

- We will consider an unlearned convolutional feature extractor from "**A mathematical theory of deep convolutional neural networks for feature extraction**" by Wiatowski and Bölcskei.
- Let $\Omega = ((\Psi_k, M_k, P_k))_{k\in\mathbb{N}}$ be a sequence where
  - $\Psi_k = \{\gamma_{\lambda_k}\}_{\lambda_k\in\Lambda_k}$, where Λ_k is an index set, is a pre-chosen filter bank (convolutional filters).
  - The operator M_k is a lipschitz nonlinearity.
  - Each P_k is a pooling operator $f \mapsto S_k^{n/2} P_k(f)(S_k\cdot)$ for $f \in L^2(\mathbb{R}^n)$ and $S_k \geq 1$ is a subsampling factor.
- Define the operators

$$U_k[\lambda_k]f := S_k^{n/2} P_k(M_k(f * g_{\lambda_k}))(S_k\cdot). \tag{1}$$

- For a path $q = (\lambda_1, \dots, \lambda_k)$, we define

$$U_k[q]f = U_k[\lambda_k]\cdots U_1[\lambda_1]f, \tag{2}$$

  where $U[\emptyset] = f$.
- Define

$$\Phi_\Omega^k(f) := \{U[q]f * \chi_k\}_{q\in\Lambda_1^k}, \tag{3}$$

  where $\chi_{k-1} = g_{\lambda_k^*}$ with $q \in \Lambda_1^k := ((\Lambda_1 \setminus \{\lambda_1^*\}) \times \cdots \times (\Lambda_k \setminus \{\lambda_k^*\})$. The feature extractor maps f to feature vector defined by $\Phi_\Omega(f) := \bigcup_{k=0}^{\infty} \Phi_\Omega^k(f)$.
- The norm/energy is measured by: $|||\Phi_\Omega(f)|||^2 := \sum_{k=0}^{\infty} \sum_{q\in\Lambda^k} \|U[q]f * \chi_k\|_2^2$.

Stability in Machine Learning

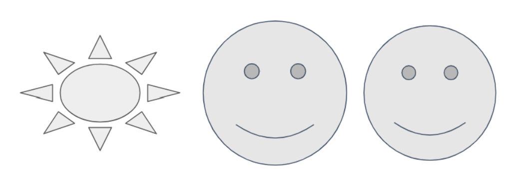


Figure: Sun or smiley face?

- Let $\Phi : \mathcal{H}_1 \to \mathcal{H}_2$ be some representation (i.e. a convolutional neural network), where $\mathcal{H}_1, \mathcal{H}_2$ are Hilbert Spaces.
- Define $L_\tau f(x) = f(x - \tau(x))$, where $\tau \in C^2(\mathbb{R}^n)$ and τ is "small."
- It would be ideal for a representation to satisfy

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

  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.

Time-Frequency Deformations

- Remember that were concerned with $L_\tau f(x) = f(x - \tau(x))$.
- When $f \in \mathbf{L}^2(\mathbb{R}^n)$, via Fourier inversion, one has

$$(L_\tau f)(x) = \int_{\mathbb{R}^n} e^{i\langle \xi, x - \tau(x)\rangle} \hat{f}(\xi)\, d\xi. \tag{4}$$

- General form of a Fourier Integral Operator:

$$\int_{\mathbb{R}^n} e^{i\Psi(x,\xi)} a(x,\xi) \hat{f}(\xi)\, d\xi. \tag{5}$$

- This motivates us to consider a deformation of the following form:

$$K_{\tau_1,b} f(x) := \int_{\mathbb{R}^n} e^{i\langle \xi, x - \tau_1(x)\rangle} \underbrace{(1 + b(x,\xi))}_{\color{red}{AmplitudeDeformation}} \hat{f}(\xi)\, d\xi \tag{6}$$

with $\tau_1 \in C^2(\mathbb{R}^n) \cap \mathbf{L}^\infty(\mathbb{R}^n)$.

Time-Frequency Deformations - Continued

- For this specific paper, we consider a separable deformation:

$$K_{\tau_1,\tau_2,\tau_3} f(x) := \int_{\mathbb{R}^n} e^{i\langle \xi, x-\tau_1(x)\rangle}(1+\tau_2(\xi)\tau_3(x))\hat{f}(\xi)\, d\xi \tag{7}$$

  with $\tau_1 \in C^2(\mathbb{R}^n) \cap \mathbf{L}^\infty(\mathbb{R}^n)$ and $\|\nabla\tau_1\|_\infty < \frac{1}{2n}$.
- Why consider time-frequency deformations in the first place?
  - Small deformations in data always occur (time-deformation).
  - Many adversarial attack methods target frequency bands, so it's important to get some measure of how robust networks will be.
  - A bound of the form
    $$\|\Phi f - \Phi K_{\tau_1,\tau_2,\tau_3} f\|_{\mathcal{H}_2} \leq G(\tau_1,\tau_2,\tau_3)\|f\|_{\mathcal{H}_1}$$
    with $G(\tau_1,\tau_2,\tau_3)$ getting smaller as the deformations get smaller would show convolutional architectures have good inductive bias against time-frequency deformations.
  - A bound like above could guide convolutional neural network design (future work).

Time-Frequency Deformation Bounds (Main Result)

- Reminder:
  - $\Phi_\Omega(f) := \bigcup_{k=0}^{\infty} \Phi_\Omega^k(f)$
  - $|||\Phi_\Omega(f)|||^2 := \sum_{k=0}^{\infty} \sum_{q \in \Lambda^k} \|U[q]f * \chi_k\|_2^2.$
  - $K_{\tau_1,\tau_2,\tau_3} f(x) := \int_{\mathbb{R}^n} e^{i\langle \xi, x-\tau_1(x)\rangle}(1+\tau_2(\xi)\tau_3(x))\hat{f}(\xi)\, d\xi$
- Define the phase shift operator $M_\omega f(x) := e^{2\pi i \omega(x)} f(x)$.
- Suppose $\hat{f}$ is supported in $B(0,R)$ for some $R > 0$, $\tau_2 \in \mathbf{L}^1(\mathbb{R}^n) \cap \mathbf{L}^\infty(\mathbb{R}^n) \cap C(\mathbb{R}^n)$, and $\tau_3 \in \mathbf{L}^\infty(\mathbb{R}^n)$.
- Then
$$|||\Phi_\Omega(M_\omega K_{\tau_1,\tau_2,\tau_3} f) - \Phi_\Omega(f)||| \leq S(\tau_1,\tau_2,\tau_3,\omega)\|f\|_2$$
  with
$$S(\tau_1,\tau_2,\tau_3,\omega) = C_1(R\|\tau_1\|_\infty + \|\omega\|_\infty) + \sqrt{2}\|\tau_3\|_\infty\|\tau_2\|_\infty.$$
- Interpretation of result
  - A smaller deformation in our signal results in a smaller deformation of our representation!
  - Our bound depends on the architecture chosen rather than the specific filter choice.

Conclusions and Future Work

- We've introduced time-frequency deformations and proven a time-frequency deformations bound for deep convolutional architectures.
- The general case where

$$K_{\tau_1,b}f(x) := \int_{\mathbb{R}^n} e^{i\langle \xi, x-\tau_1(x)\rangle}(1+b(x,\xi))\hat{f}(\xi)\,d\xi$$

  is still unsolved.
- It would be interesting to extend these ideas to more general convolutional architectures.
- Could time-frequency deformations be used as an adversarial attack?
  - See "ADef: an Iterative Algorithm to Construct Adversarial Deformations" for small time deformation attacks.
  - How does training an architecture with random time-frequency deformations as augmentation affect adversarial robustness of the model?

References

1. T. Wiatowski and H. Bölcskei, "A mathematical theory of deep convolutional neural networks for feature extraction," IEEE Transactions on Information Theory, vol. 64, no. 3, pp. 1845–1866, 2017.
2. Alaifari, Rima, Giovanni S. Alberti, and Tandri Gauksson. "ADef: an Iterative Algorithm to Construct Adversarial Deformations." International Conference on Learning Representations.