TTIC 31230, Fundamentals of Deep Learning

David McAllester, April 2017

Some Linear Systems and Wavelet Theory

Linear Systems Theory

An amplifier takes in a low power signal (your favorite tune) and scales it up to drive the speakers.

A good amplifier is linear — $A(\alpha x + \beta y) = \alpha A(x) + \beta A(y)$.

Deviation from linearity is called distortion.

Observation: Any linear shift-invariant operator on signals is completely characterized by its impulse response — how it responds an impulse input.

Theorem: Any linear shift-invariant operator on signals is completely characterized by its frequency response — how it responds to a pure tone.

Theorem: The frequency response is the Fourier transform of the impulse response.

Classical Convolution

Let $x(t)$ be the input signal and let $f(\tau)$ be the impulse response of an operator (or amplifier) A.

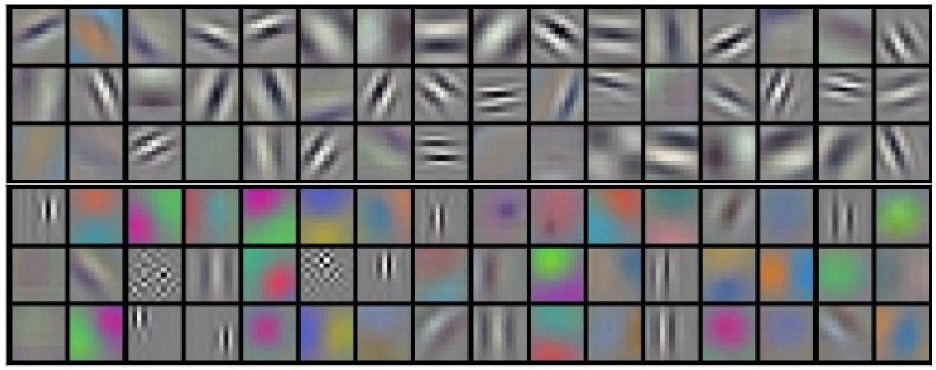
For a causal circuit (it does not response before it is stimulated) with an impulse response of finite duration T, the output $A(x)$ can be written as follows.

$$A(x)(t) = \int_{\tau=0}^{T} x(t-\tau)f(\tau)d\tau$$

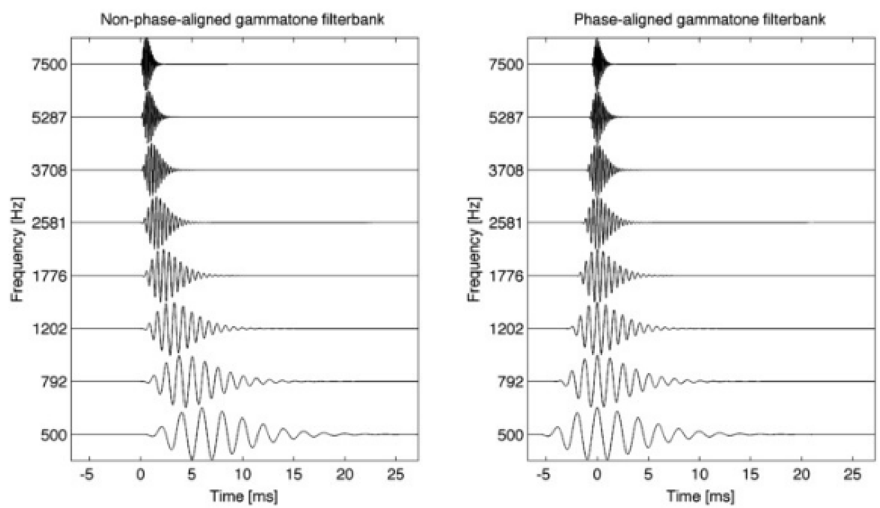
This is the standard definition of convolution $A(x) = x * f$ (for causal bounded f).

Surprisingly, convolution is commutative and we have $A(x) = f * x$ where f is viewed as a finite duration signal and x is viewed as an unbounded a-causal impulse response.

Why Are Early Filters Wavelets?



From Krizhevsky



from MathWorks

Invariance

In machine learning it is standard to assume a data distribution — a probability distribution over problem instances from which a training set has been sampled.

Consider the distribution of "natural" 360 degree images.

It should be true that the probability of a given 360 degree image is equal to the probability of any rotation of that image.

We say that the probability distribution is invariant to rotation.

Invariance theory is more significantly applied to local image patches.

Invariances

Translation invariance (stationarity)

Scale invariance in images.

Time-compression invariance in audio.

Rotation invariance for image patch distributions.

PCA and Invariance

Principal component analysis (PCA) is a dimensionality reduction method in which a signal (vector, image, or sound wave) is represented by its projection onto "principal component signals".

The principal components are the eigenvectors of the covariance matrix.

The principal components of the covariance matrix of a stationary signal distribution (a stationary signal process) are the Fourier basis signals.

PCA and Invariance

The eigenvalues of the covariance matrix are given by the power spectrum of the signal distribution.

This is the Einstein-Wiener-Khinchin theorem (proved by Wiener, and independently by Khinchin, in the early 1930s, but — as only recently recognized – stated by Einstein in 1914). From "Signals and Systems" by Oppenheim and Verghese

This explains projection onto complex exponentials as a first step in signal processing and signal compression (e.g., JPEG).

More Formally

Let ρ be a probability density over vectors in $\mathbb{R}^n$.

We say ρ is rotation-stationary if

- $\mathrm{E}[x_i] = \mathrm{E}\left[x_j\right]$ for all i, j.
- $\mathrm{E}\left[x_i x_j\right] = f(i - j \bmod n)$

Rotation stationarity is a simplification of the more widely used notion of translation stationarity (or just stationarity).

More Formally

The covariance matrix is given by

$$\Sigma_{i,j} = \mathrm{E}\left[x_i x_j - \mathrm{E}\left[x_i\right]\mathrm{E}\left[x_j\right]\right] = g(i - j \bmod n)$$

A matrix satisfying $\Sigma_{i,j} = g(i-j \bmod n)$ is called **circulant**.

The eigenvectors of a circulant matrix form a discrete Fourier basis.

Invariance and Data Augmentation

CNNs build translation invariance into the architecture.

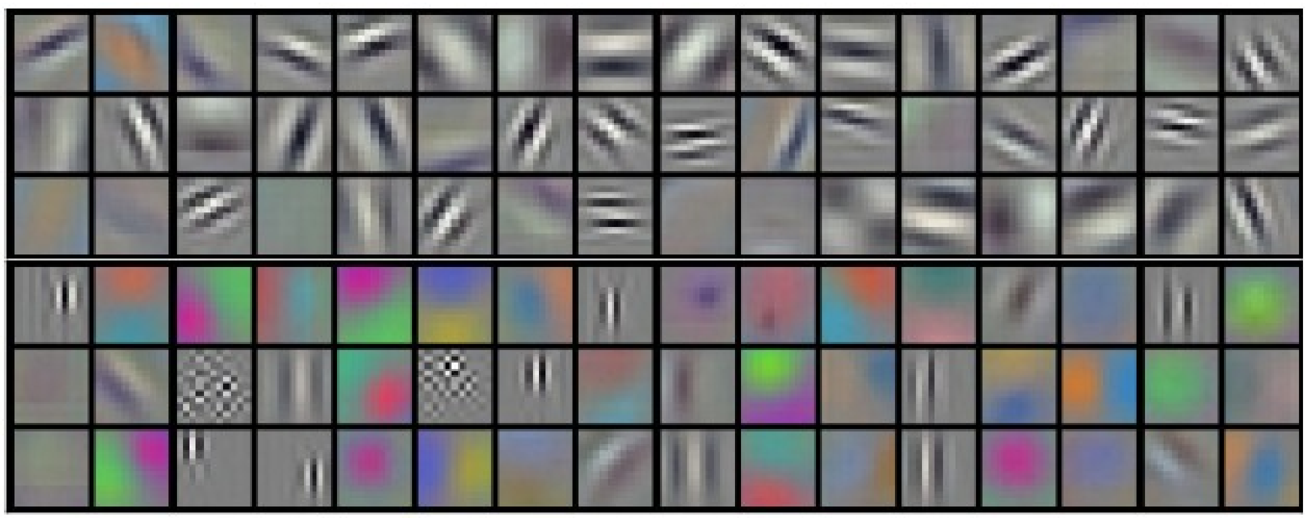
Another approach to invariance is to apply invariant transformations to the training data.

For example we can apply translations, scalings, rotations, reflections to generate more labeled images in MNIST or Imagenet to get a much larger training set.

Wavelets

In practice we want the compressed representation to be local to also satisfy scale invariance. This leads to **wavelets**.

To my knowledge scale invariance is not currently built into deep vision architectures. (But talk to Michael Maire at TTIC.)



END