0.1 Properties of a Scatternet

Adversarial attacks are of the form of additive noise. One of the nice properties of Scattering transforms is that

$$\|\Phi(x) - \Phi(x+\epsilon)\| \leq \epsilon$$

Other transformations such as camera warping come under the scope of diffeomorphisms:

$$L_\tau x(u) = x(u - \tau(u))$$

We can define the largest displacement of this field as

$$\|\tau\|_\infty = \sup_{u \in \mathbb{R}^2} |\tau(u)|$$

Deformations not only change u, but they change x as well. A Taylor series expansion around u shows this:

$$u + v - \tau(u+v) \approx u + v - \nabla\tau(u)v = u - \tau(u) + (1 - \nabla\tau(u))v$$

This can be summarised by noting that in the neighbourhood of u, τ introduces a translation by $\tau(u)$ and a warping that differes from 1 by $\nabla\tau(u)$. This warping can be quantified as

$$\|\nabla\tau\|_\infty = \sup_{u \in \mathbb{R}^2} \|\nabla\tau(u)\|$$

A representation is stable to deformations if we can define to small constants, C_1, C_2 such that for all $x \in L^2(\mathbb{R}^2$:

$$\|\Phi L_\tau x - \Phi x\| \leq (C_1 \|\tau\|_\infty + C_2 \|\nabla\tau\|_\infty) \|x\|$$

Note that if $C_1 = 0$ then we have full translation invariance (translation is when $\tau(u) = C$, $\nabla\tau(u) = 0$). Full translation invariance does not imply stability to transformations. E.g. the Fourier modulus has full translation invariance. If we introduce a warping:

$$\tau(u) = \epsilon u, \epsilon > 0$$

Then a sine wave with frequency ω will get shifted to $\frac{w}{1-\epsilon}$ and $\|\Phi L_\tau x - \Phi x\| = 2$ even when $\|\nabla\tau\|_\infty = \epsilon$ is made arbitrarily small.

Is it possible to maintain these properties if I modify the Scattering Transform? So the translation invariance property isn't really invariance. Well it is is invariant to sub-pixel and to an extent pixel shifts.

0.1.1 Nonlinearities

The wavelet operator

$$W[\lambda]x = x * \psi_\lambda$$

commutes with translations but

$$\int W[\lambda]x(u)du = 0$$

because $\int \psi(u)du = 0$. To get a non-zero invariant, need to 'demodulate', mapping $W[\lambda]x$ to a lower frequency with a non-zero integral. Recall a simple Morlet wavelet has form:

$$\psi(u) = e^{j\eta u}\phi(u)$$

(first term is the modulation and second term is low pass). Then

$$\psi_\lambda(u) = e^{j\lambda\eta u}\phi_\lambda(u)$$

and

$$W[\lambda]x(u) = e^{j\lambda\eta u}\left(x^\lambda * \phi_\lambda(u)\right)$$

with $x^\lambda(u) = e^{-j\lambda\eta u}x(u)$. A simple non-linearity would be to cancel the wavelet and the signals modulating term, i.e.

$$M[\lambda]h(x) = e^{-j\lambda\eta u}e^{-j\Phi(\hat{h}(\lambda\eta))}h(x)$$

where $\Phi(\hat{h}(\lambda\eta))$ is the complex phase of $\hat{h}(\lambda\eta)$. Then

$$\begin{aligned} M[\lambda]W[\lambda]x(u)du &= \int e^{-j\lambda\eta u}e^{-j\Phi(\hat{h}(\lambda\eta))}\left(e^{j\lambda\eta u}\left(x^\lambda * \Phi_\lambda(u)\right)\right)du \\ &= \int e^{-j\Phi(\hat{h}(\lambda\eta))}\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{j\Phi(\hat{h}(\lambda\eta)}|\hat{x}(\lambda\eta)||\hat{\Phi}(0)|e^{2\pi j\omega u}d\omega du \\ &= |\hat{x}(\lambda\eta)||\hat{\Phi}(0)| \end{aligned}$$

This just gives us the the fourier modulus, which we saw earlier was a poor choice as it is not stable to diffeomorphisms. This implies that demodulation is not the greatest thing to do.

0.2 Enter DCFNet

This was an idea I'd had as well, but [1] does a good job at formalizing the properties. The idea is to compose a regular CNN filter as:

$$h_f^{(l)}(c,\mathbf{u})=\sum_{k=1}^{K}a_f(c,k)\psi_k(\mathbf{u})$$

where $h_f^{(l)}(c,\mathbf{u})$ is the f-th filter in the l-th layer with channel coordinate c and spatial coordinates $\mathbf{u}$, ψ_k are predefined basis functions and $a_f(c,k)$ are the learned expansion coefficients combining the k different bases for each input channel.

Consider a spatial deformation denoted by $D_\tau:\mathbb{R}^2\to\mathbb{R}^2$ given by:

$$D_\tau x(c,\mathbf{u})=x(c,\mathbf{u}-\tau(\mathbf{u}))=x(c,\rho(\mathbf{u})),\quad \forall\mathbf{u},c$$

recall that c indexes the channel domain, so we are assuming that the same diffeomorphism applies to all channels equally. Assume that the distortion is bounded, specifically:

$$|\nabla\tau|_\infty=\sup_u\|\nabla\tau(u)\|<C$$

The boundedness implies ρ^{-1} exists locally. We want to control

$$\left\|x^{(L)}\left[D_\tau x^{(0)}\right]-x^{(L)}\left[x^{(0)}\right]\right\|$$

so that when the input undergoes a deformation the output at the L-th layer is not severely changed. They show in their network that $\left\|x^{(L)}\left[D_\tau x^{(0)}\right]-x^{(L)}\left[x^{(0)}\right]\right\|$ is bounded by the magnitude of the deformation up to a constant proportional to the norm of the signal.

Define the L^1 and L^2 norms and the average energy of $x(c,\mathbf{u})$ to be:

$$\|x\|_1 = \sum_{c=1}^{C}\int_{\mathbb{R}^2}|x(c,\mathbf{u})|d\mathbf{u} \tag{0.2.1}$$

$$\|x\|_2^2 = \sum_{c=1}^{C}\int_{\mathbb{R}^2}|x(c,\mathbf{u})|^2 d\mathbf{u} \tag{0.2.2}$$

$$\|x\|_{av}^2 = \frac{1}{C|\Omega|}\|x\|^2 \tag{0.2.3}$$

Let the number of channels at layer l be M_l, the largest filter norm is:

$$A_l = \sup_f \sum_{c=1}^{M_{l-1}} \left\| h_f^{(l)}(c, \mathbf{u}) \right\|_1 \tag{0.2.4}$$

$$B_l = \sup_c \frac{M_{l-1}}{M_l} \sum_{f=1}^{M_l} \left\| h_f^{(l)}(c, \mathbf{u}) \right\|_1 \tag{0.2.5}$$

$$C_l = \max\{A_l, B_l\} \tag{0.2.6}$$

Consider the largest filter norm over the f filters at layer l:

$f : X \to Y$ is Lipschitz continuous if there exists a real constant $K \geq 0$ such that for all $x_1, x_2 \in X$:

$$d_Y(f(x_1), f(x_2)) \leq K d_X(x_1, x_2)$$

It is clear that the complex magnitude is Lipschitz continuous with constant $K = 1$ as:

$$\begin{aligned} d_Y(f(x_1), f(x_2)) &= ||w| - |z|| \\ &\leq |w - z| \\ &= d_X(x_1, x_2) \end{aligned}$$

Where the second line holds by the reverse triangle inequality.

References

[1] Q. Qiu, X. Cheng, R. Calderbank, and G. Sapiro, “DCFNet: Deep Neural Network with Decomposed Convolutional Filters”, *arXiv:1802.04145 [cs, stat]*, Feb. 2018. arXiv: 1802.04145 `[cs, stat]`.