The Theory of the Scattering Transform

Edouard Oyallon
edouard.oyallon@cnrs.fr
CNRS, ISIR

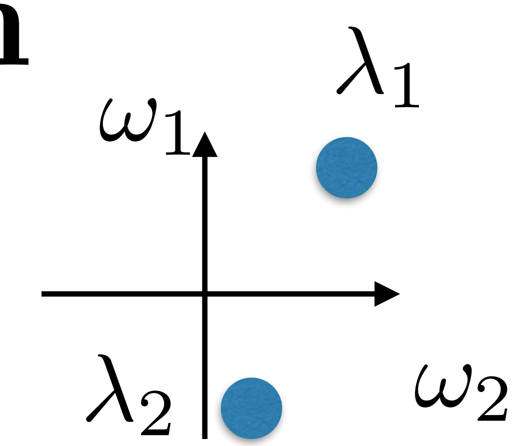
cnrs

Machine Learning & Deep Learning for Information Access

Announcement, reminders

- Homework, group projects are due next Friday.
  I will apply penalty to late groups. . .

Definition of the Scattering Transform



Define a path of length n as $(\lambda_1, ..., \lambda_n)$ where $\lambda = (\theta, 2^{-j}), |\lambda| = 2^{-j}$

Let us fix mother wavelet ψ and low-pass filter ϕ, smooth, with fast decay.

Definition: The Scattering path of $S_J[\lambda_1, ..., \lambda_n]x$ is given by:

$$S_J[\lambda_1, ..., \lambda_n]x \triangleq ||...|x \star \psi_{\lambda_1}| \star ...| \star \psi_{\lambda_n}| \star \phi_J$$

Definition: Scattering Transform of order n:

$$S_J^n x \triangleq \{S_J[\lambda_1, ..., \lambda_k]x\}_{\lambda_1, ..., \lambda_k, k \leq n}$$

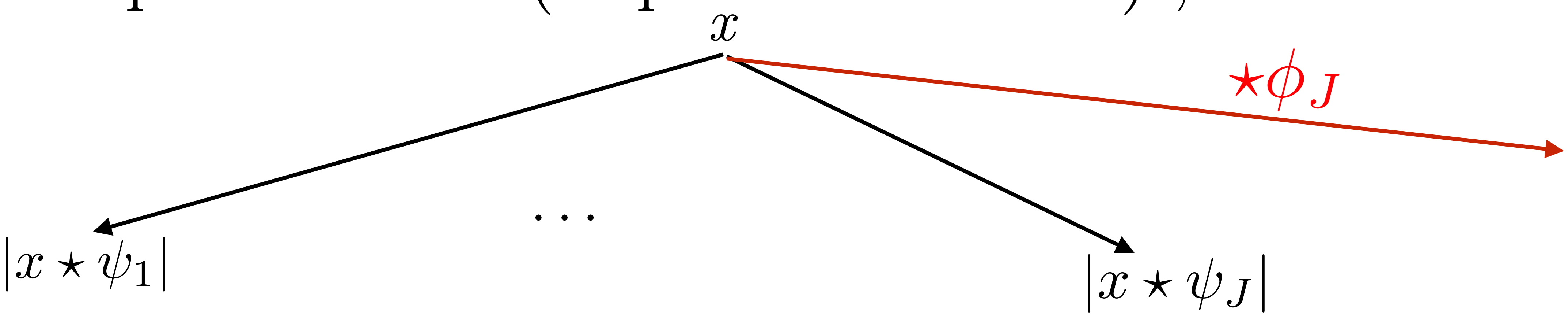
Its norm si given by:

$$\|S_J^n x\|^2 = \sum_{\lambda_1, ..., \lambda_k, k \leq n} \|S_J[\lambda_1, ..., \lambda_k]x\|^2$$

We will also write the Scattering Transform as: $S_J x \triangleq \{S_J^n x\}_{n \geq 0}$

The Scattering Transform

- Main principle: cascade wavelets AND modulus non-linearity. Depth: "order" (in practice order 2) ; J: "Scale"



$$S_J x = \{x \star \phi_J,$$

Scattering Transform Theory

Useful technical lemma

Technical Lemma

Lemma (Schur Lemma): Let $Kf(u) = \int_v k(u,v) f(v)\, dv$

$$\int_u |k(u,v)|\, du \leq C_1 \text{ and } \int_v |k(u,v)|\, dv \leq C_2 \Rightarrow \|K\| \leq \sqrt{C_1 C_2}$$

Lemma (Non-expansivity): for $x, y \in L^2(\mathbb{R}^d)$ then $\| |x| - |y| \| \leq \|x - y\|$

Lemma (Diffeomorphisms): If $\|\nabla\tau\|_\infty \leq \frac{1}{2}$ then: $\|L_\tau\| \leq 2^d$

and
$$|1 - \det(\mathbf{I} - \nabla\psi(u))| \leq d\|\nabla\tau\|_\infty$$
$$2^{-d} \leq |\det(\mathbf{I} - \nabla\tau)(u)| \leq 2^d$$

Lemma (Derivation, Fourier): $\psi \in L^1(\mathbb{R}^d) \Rightarrow \hat{\psi} \in L^\infty(\mathbb{R}^d)$

and $\psi \in L^1(\mathbb{R}^d), (u \rightarrow \|u\|\psi) \in L^1(\mathbb{R}^d) \Rightarrow \hat{\psi} \in \mathcal{C}^1$

ψ is $\mathcal{C}^1, \psi \in L^1(\mathbb{R}^d), \|\nabla\psi\| \in L^1(\mathbb{R}^d) \Rightarrow (\omega \rightarrow \|\omega\|\hat{\psi}(\omega)) \in L^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$

Lemma (Littlewood Paley): If ψ is $\mathcal{C}^1$ and $\int \psi = 0$ and if

$$\max(\|\nabla\psi(u)\|, |\psi(u)|) \leq C \frac{1}{1 + \|u\|^{2(d+1)}}$$

then, there is C' such that:

$$\forall \omega \in \mathbb{R}^d, \sum_{j \geq 0} |\hat{\psi}(2^j \omega)|^2 \leq C'$$

Intermediary notations

$$A_J x = x \star \phi_J \text{ and } V_J x = \{x \star \psi_\lambda\}_{\lambda \in \Lambda}$$

s.t.: $W_J x = V_J x \cup A_J x$

and:

$$U_n[\lambda_1, \ldots \lambda_n]x \triangleq |\ldots|x \star \psi_{\lambda_1}|\ldots \star \psi_{\lambda_n}|$$

or

$$U_{n+1}x = |V_j U_n x|$$

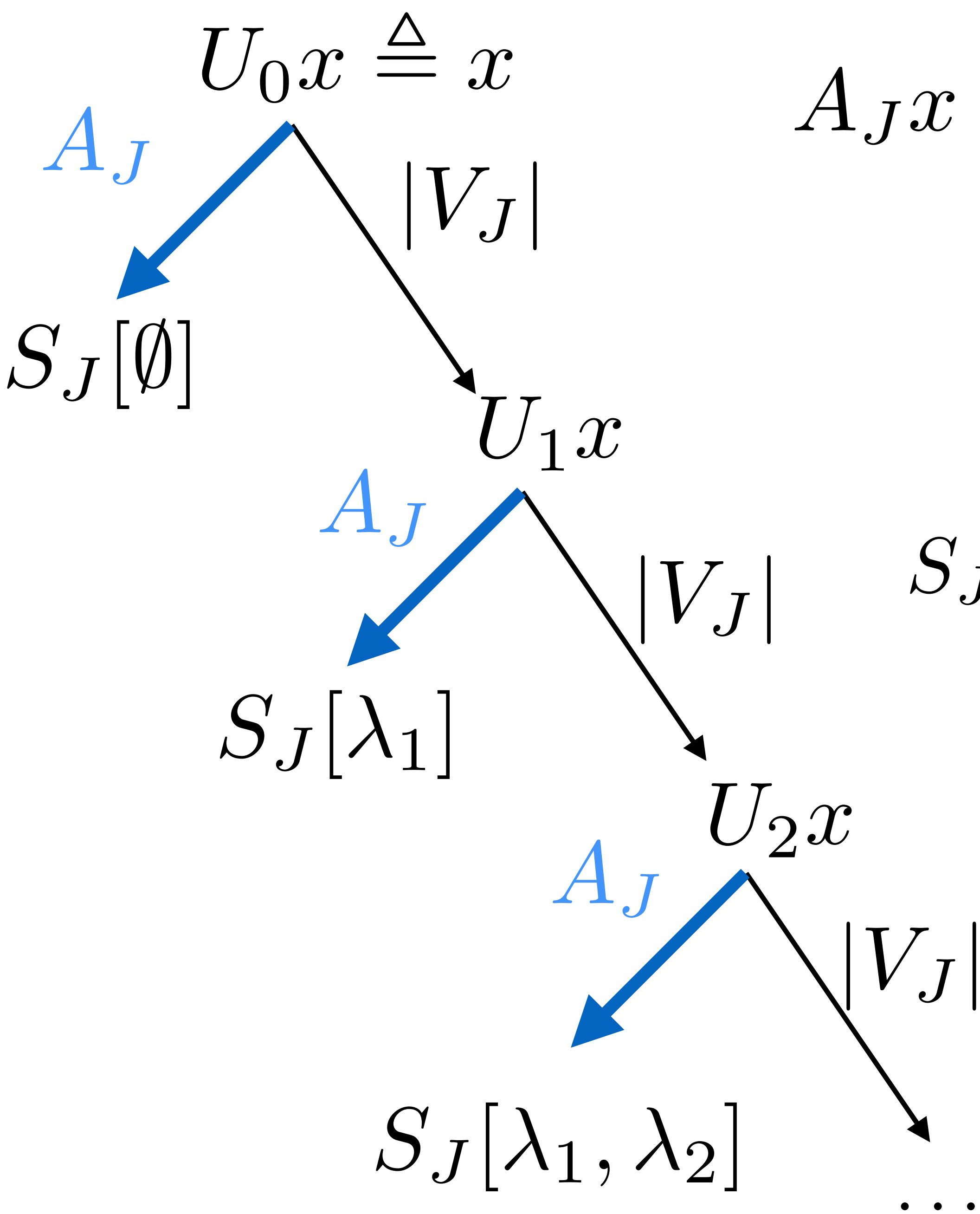
← point-wise modulus

We can introduce here an intermediary norm:

$$\|U_n x\|^2 = \sum_{(\lambda_1, \ldots, \lambda_n)} \|U_n[\lambda_1, \ldots \lambda_n]x\|^2$$

Scattering Transform defined via Integral Operators

Coefficients of the scattering transform are given by:



$$A_J x = x \star \phi_J \text{ and } V_J x = \{x \star \psi_\lambda\}_{\lambda \in \Lambda}$$

$$\text{s.t.: } W_J x = V_J x \cup A_J x$$

Here:

$$S_J[\lambda_1, ..., \lambda_n]x = ||...|x \star \psi_{\lambda_1}| \star ...| \star \psi_{\lambda_n}| \star \phi_J$$
$$= A_J U_n[\lambda_1, ..., \lambda_n]x$$

Scattering of order n:

$$S_J^n x = \{\cup_{\lambda_1, ..., \lambda_j \in \Lambda, j \leq n} S_J[\lambda_1, ..., \lambda_j]\}$$

Definition & non-expansivity

Non-expansivity of Scattering Transform

- Proposition: Given $x \in L^2(\mathbb{R}^d), y \in L^2(\mathbb{R}^d)$ we have, if $\|W_J\| \leq 1$:

$$\|S_J x - S_J y\| \leq \|x - y\|$$

Proof:

Lemma: for $x, y \in L^2(\mathbb{R}^d)$:

$$\||x| - |y|\| \leq \|x - y\|$$

Remark: implies existence

Stability to deformations

Main theorem statement

Recall that: $W_J x = \{x \star \psi_{j,\theta}, x \star \phi_J\}_{\theta, j \leq J}$

- Theorem (Adapted from Mallat, 2012): If ϕ, ψ are regular enough, $\|W_J\| \leq 1$ and if $\int_u \psi(u)\,du = 0$, there exists C such that for any J, if $\|\nabla\tau\|_\infty \leq \frac{1}{2}$, then:

$$\|S_J^n L_\tau x - S_J^n x\| \leq n^{3/2} C \|x\| (\|\nabla\tau\|_\infty + \|\Delta\tau\|_\infty + \frac{\|\tau\|_\infty}{2^J})$$

In other words, the Scattering Transform is stable to small deformations.

Typical applications: n=2, J=3

Sketch of the proof.

Write: $[A,B] = AB - BA$ which measures how A, B commute.

- First, we note that:

$$\|S_J^n x - S_J^n L_\tau x\|^2 = \sum_{m\leq n} \|A_J U_m x - A_J U_m L_\tau x\|^2$$

- Next, we will bound each Scattering "paths"

$$\|A_J U_n x - A_J U_n L_\tau x\| \leq \left(c_1 \|A_J L_\tau - A_J\| + n c_2 \|[L_\tau, V_j]\|\right) \|x\|$$

- Finally, we will bound each operators:

$$\|A_J L_\tau - A_J\| \leq C_1 (2^{-J} \|\tau\|_\infty + \|\nabla \tau\|_\infty)$$

and

$$\|[L_\tau, V_J]\| \leq C_2 (\|\nabla \tau\|_\infty + \|\Delta \tau\|_\infty)$$

- In conclusion:

$$\|S_J^n L_\tau x - S_J^n x\| \leq n^{3/2} C \|x\| (\|\nabla \tau\|_\infty + \|\Delta \tau\|_\infty + \frac{\|\tau\|_\infty}{2^J})$$

Proof step 1

$$A_J x = x \star \phi_J \text{ and } V_J x = \{x \star \psi_\lambda\}_{\lambda \in \Lambda}$$
$$\text{and } W_J = \{A_J, V_J\}$$

- Assume we proved that for ϕ, ψ regular enough, we get:

$$\|A_J L_\tau - A_J\| \leq C_1 (2^{-J} \|\tau\|_\infty + \|\nabla \tau\|_\infty)$$
and
$$\|[L_\tau, V_J]\| \leq C_2 (\|\nabla \tau\|_\infty + \|\Delta \tau\|_\infty)$$

- <u>Theorem</u>: If ϕ, ψ are regular enough, $\|W_J\| \leq 1$ and if $\int_u \psi(u)\, du = 0$, there exists C such that for any J, if $\|\nabla \tau\|_\infty \leq \frac{1}{2}$, then:

$$\|S_J^n L_\tau x - S_J^n x\| \leq n^{3/2} C \|x\| (\|\nabla \tau\|_\infty + \|\Delta \tau\|_\infty + \frac{\|\tau\|_\infty}{2^J})$$

constants are suboptimal

Stability to deformations: low-pass filter

- Proposition: Assume: $\int_u \|\nabla\phi(u)\|\, du < \infty$ and $\int_u |\phi(u)|\, du < \infty$

  Then there exists C>0 such that for any J and $\|\nabla\tau\|_\infty \leq \frac{1}{2}$:

$$\|A_J - A_J L_\tau\| \leq C(2^{-J}\|\tau\|_\infty + \|\nabla\tau\|_\infty)$$

Commutation with deformations of high-frequencies

- Proposition: Assume that ψ is regular and $\int_u \psi(u)\,du = 0$ then there exists C such that for any J and $\|\nabla\tau\|_\infty \leq \frac{1}{2}$:

$$\|[L_\tau, V_J]\| \leq C(\|\nabla\tau\|_\infty + \|\Delta\tau\|_\infty)$$

Summary of the Scattering's properties we discussed

- Scattering is stable:
$$\|S_J x - S_J y\| \leq \|x - y\|$$
- Linearize small deformations:
$$\|S_J L_\tau x - S_J x\| \leq C \|\nabla \tau\| \|x\|$$
- Invariant to local translation:
$$\|a\| \ll 2^K \Rightarrow S_J L_a x \approx S_J x$$
- For λ, u, $S_J x(u, \lambda)$ is **covariant** with :

$$\text{if } \forall u \forall g \in SO_2(\mathbb{R}), g.x(u) \triangleq x(g^{-1}u)$$

$$S_J(g.x)(u, \lambda) = S_J x(g^{-1}u, g^{-1}\lambda) \triangleq g.S_J x(u, \lambda)$$

More Scattering

Scattering moments

Ref.: Invariant Convolutional Scattering Network, J. Bruna and S Mallat

- For a stationary process X (e.g., a texture)

$$E(X \star f) = E(X) \star f$$

- This leads to the Expected Scattering:

$$\bar{S}[\lambda_1] = \mathbb{E}|X \star \psi_{\lambda_1}|$$

Modulus is important as expectation would be 0!

$$\bar{S}[\lambda_1, \lambda_2] = \mathbb{E}||X \star \psi_{\lambda_1}| \star \psi_{\lambda_2}|$$

can be estimated via an unbiased estimator:

$$S[\lambda_1, \lambda_2]X = \int ||X \star \psi_{\lambda_1}| \star \psi_{\lambda_2}|$$

Roto-translation scattering

Ref.: PhD of L. Sifre

$$S_0 x = \int_u x(u)\,du \quad \text{and} \quad Y^1_{j_1}(u,\theta_1) = |x \star \psi_{j_1,\theta_1}(u)|$$

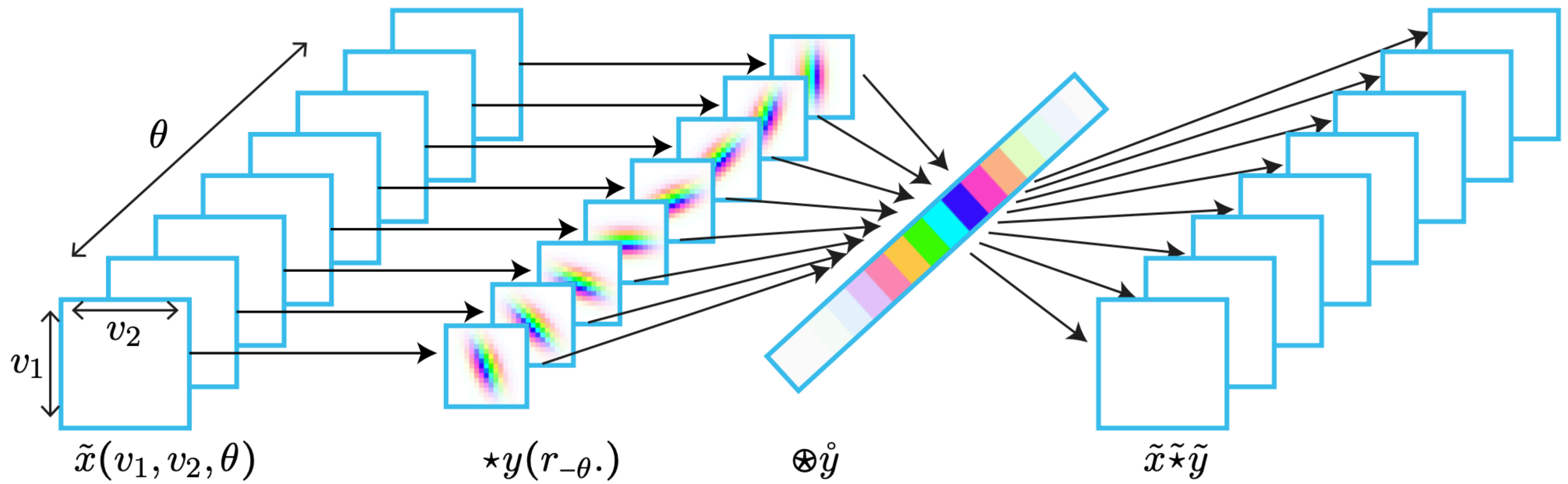
Let $S_1 x = \int_{u,\theta} Y^1(u,\theta)\,du\,d\theta$ and $\Psi(u,\theta) = \psi_{j_2,\theta_2}(u)\psi_k(\theta)$

then, we get:

$$Y^2_{j_1,j_2,\theta_2,k}(\theta,u) = \int_{\theta',u'} |x \star \psi_{j_1,\theta'}(u')|\psi_{j_2,\theta_2+\theta'}(u-u')\psi_k(\theta-\theta')du\,d\theta$$

Let $S_2 x = \int_{u,\theta} Y^2(u,\theta)\,du\,d\theta$

- Then Sx is invariant to roto-translation.

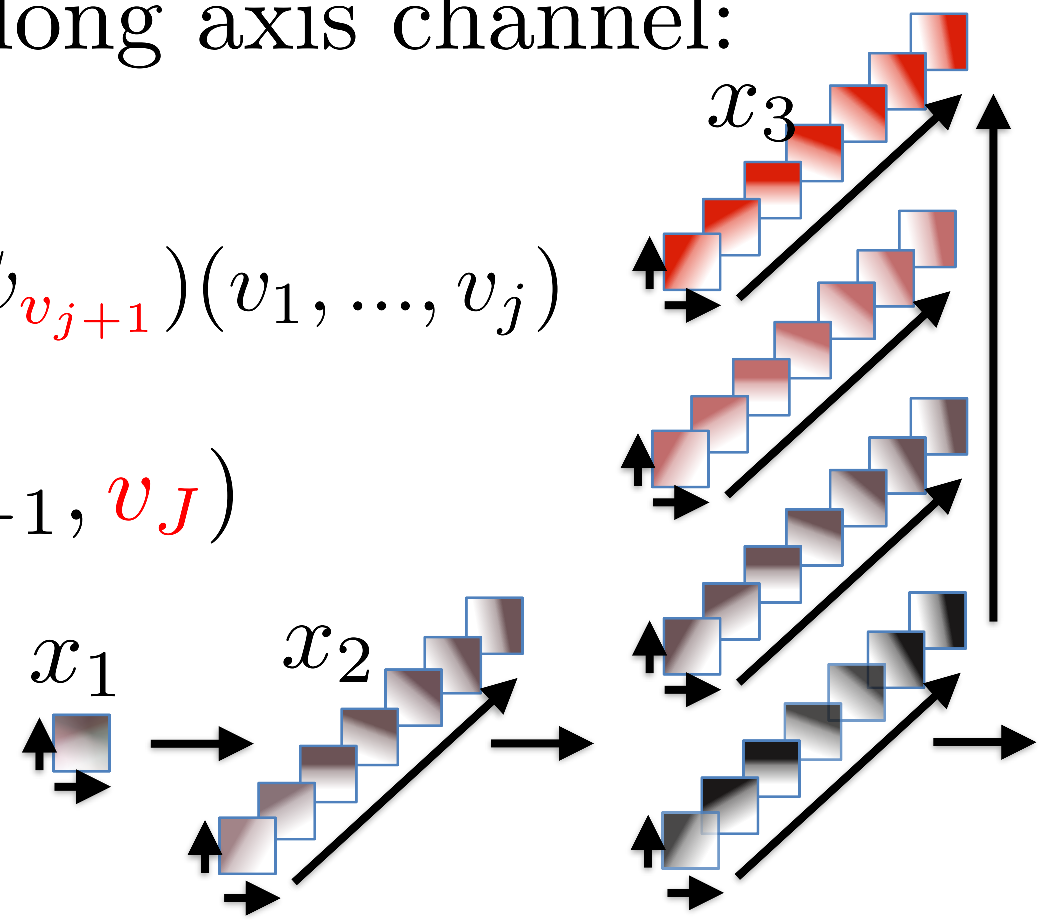


One generalization among many

- CNN that is convolutional along axis channel:

$$x_{j+1}(v_1, ..., v_j, {\color{red}v_{j+1}}) = \rho_j(x_j \star^{v_1,...,v_j} \psi_{\color{red}v_{j+1}})(v_1, ..., v_j)$$

$$x_J({\color{red}v_J}) = \sum_{v_1,...,v_{J-1}} x_{J-1}(v_1, ..., v_{J-1}, {\color{red}v_J})$$



Ref.: Hiearchical CNNs, Jacobsen et al.

- For x_j, we refer to the variable v_j as an attribute that discriminates previously obtained layer.

- Representation is finally averaged: invariant along translations by v.

Very similar to equivariant CNNs

Today's lab

Mutual information

- Measure of the mutual dependence between two variables:

$$I(X;Y) = D_{KL}(P_{(X,Y)} || P_X \otimes P_Y)$$

- Also writes:

$$I(X;Y) = \int_{\mathcal{X}} \int_{\mathcal{Y}} p_{(X,Y)}(x,y) \log \frac{P_{(X,Y)}(x,y)}{P_X(x) P_Y(y)} \, dx \, dy$$

- For any diffeomorphisms, ψ, ϕ we have:

$$I(\psi(X), \phi(Y)) = I(X,Y)$$

Reducing mutual information (Information bottleneck)

- Reducing the information sounds relevant:

$$I(X;Y)=\int_{\mathbb{R}^2} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}\mathrm{d}x\mathrm{d}y = H(X)-H(X|Y)$$

Ref.: Opening the Black Box of Deep Neural Networks via Information, R Shwartz-Ziv and N Tishby

Measures the dependancy between variables

$$I(X;\Phi_1 X)\geq I(X;\Phi_2 X)\geq \ldots \geq I(X;\Phi_J X)$$

"Compress" X

$$I(X;Y)\geq I(\Phi_1 X;Y)\geq \ldots \geq I(\Phi_J X;Y)$$

... but "reveal" Y

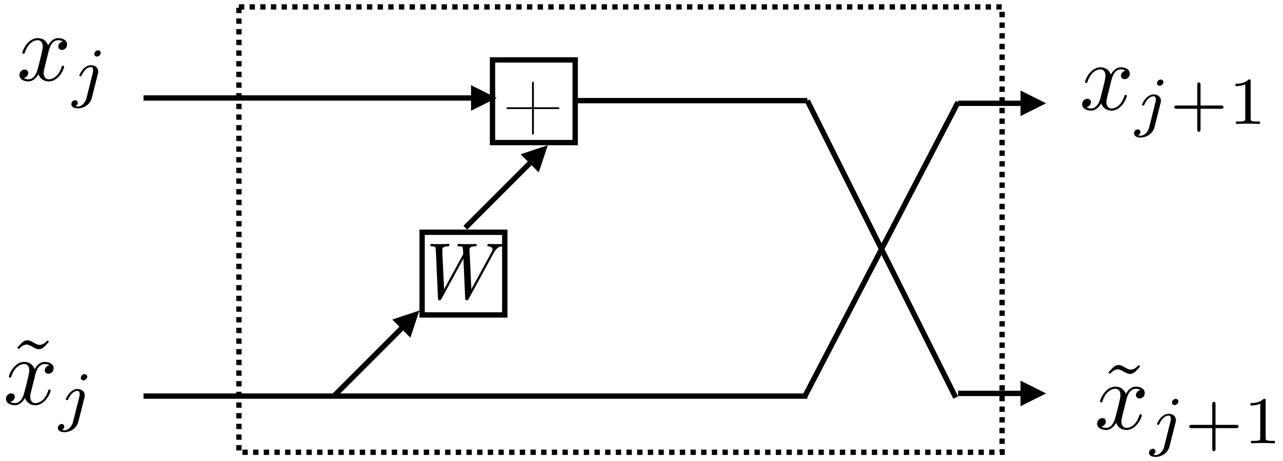
They propose to introduce:

$$\Phi_{j,\lambda} = \arg\inf_{\Phi} I(\Phi_{j-1}X,\Phi_j X)-\lambda I(\Phi_j X,Y)$$

- But one can easily build invertible CNNs...

Ref.: i-RevNet: Deep Invertible Networks, J Jacobsen et al.

$$x^t = \Phi^{-1}(t\Phi x^0 + (1-t)\Phi x^1)$$



One bijective layer