Lattice theory and algebraic models for deep learning based on mathematical morphology

Gustavo (Jesús) Angulo

Mines Paris, PSL University
CMM-Centre for Mathematical Morphology

CaLISTA Workshop – Geometry-Informed Machine Learning

Mines Paris – Paris – September 2–5 2024

"Work on deep learning or perish": *folklore wisdom*

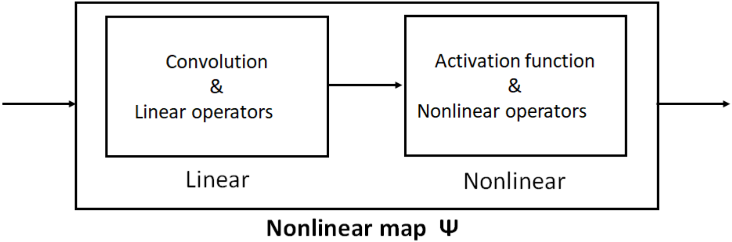
Mathematical Models "to understand" Deep Learning Neural Networks

Four recent paradigms

- KAM: Kolmogorov–Arnold Networks (Liu et al., 2024): Limited abstract representation, great numerical framework
- Categorical Deep Learning: An Algebraic Theory of Architectures (Gavranović et al., 2024): Highest abstraction
- The Real Tropical Geometry of Neural Networks (Brandenburg et al., 2024): ReLU networks and tropical geometry
- PDE-based Group Equivariant Convolutional Neural Networks (Smets et al., 2023): Networks layers as PDE-solvers

Mathematical Models "to understand" Deep Learning Neural Networks

Typical operators (layers) in a CNN block



My goal is to propose a representation theory which inspires new architectures and layers too

Plan

1. Mathematical Morphology
2. A universal representation theorem for nonlinear (increasing) operators
3. Morphological activations
4. Group Morphology and equivariant universal representation
5. Convolution as an operator in an inf-semilattice
6. Morphological model of a CNN
7. Inf-semilattice of down/up-sampling and pyramids
8. Conclusions and Perspectives

1 Mathematical Morphology
2 A universal representation theorem for nonlinear (increasing) operators
3 Morphological activations
4 Group Morphology and equivariant universal representation
5 Convolution as an operator in an inf-semilattice
6 Morphological model of a CNN
7 Inf-semilattice of down/up-sampling and pyramids
8 Conclusions and Perspectives

What is mathematical morphology? (1/2)

- An abstract algebraic formulation of the theory on complete lattices, which requires very little assumptions to be instantiated into a specific lattice structure of the space of interest
- A common representation theory of operators for the Boolean and the semicontinuous function cases, in which, for instance, any translation-invariant increasing, upper semicontinuous operator can be represented exactly as a minimal superposition of morphological erosions or dilations
- An intimate relationship with the random set theory via the notion of Choquet capacity from stochastic geometry
- Strong connections with idempotent mathematics (max-plus and max-min algebra and calculus) and tropical geometry
- Continuous models which correspond to Hamilton–Jacobi PDEs, relevant also in optics and optimal control

What is mathematical morphology? (2/2)

- A powerful extension to the case of morphology on groups, which bring a proper dealing with space symmetries and provide equivariant operators to the groups of transforms relevant in computer vision
- Multiscale operators and semigroups formulated in Riemannian, metric and ultrametric spaces
- Multiple morphological representations that provide a rich family of shape-based and geometrical descriptions and decompositions: skeletons, pattern spectra and size distributions, topological description of functions using maxima-minima extinction values, etc.
- A privileged mathematical tool for Lipschitz characterization and regularization
- A counterpart of the perceptron which yields to the scope of morphological neural networks, morphological associative memories

1. Mathematical Morphology
2. A universal representation theorem for nonlinear (increasing) operators
3. Morphological activations
4. Group Morphology and equivariant universal representation
5. Convolution as an operator in an inf-semilattice
6. Morphological model of a CNN
7. Inf-semilattice of down/up-sampling and pyramids
8. Conclusions and Perspectives

Dilation and Erosion

A lattice operator ψ is called increasing operator (or isotone) is if it is order-preserving, i.e., $\forall X, Y$

$$X \leq Y \implies \psi(X) \leq \psi(Y)$$

Dilation δ and Erosion ε are increasing operators that which satisfy

$$\delta\left(\bigvee_{i \in J} X_i\right) = \bigvee_{i \in J} \delta\left(X_i\right); \qquad \varepsilon\left(\bigwedge_{i \in J} X_i\right) = \bigwedge_{i \in J} \varepsilon\left(X_i\right)$$

Translation-invariant increasing set operators of set X by structuring element Y

- Dilation: $\delta_Y(X) = X \oplus Y = \bigcup_{y \in Y} X_y = \{x + y \ : \ x \in X, y \in Y\} = \{p \in E \ : \ X \cap \check{Y}_p \neq \emptyset\}$
- Erosion: $\varepsilon_Y(X) = X \ominus Y = \bigcap_{y \in Y} X_{-y} = \{p \in E \ : \ Y_p \subset X\} = \{x \ : \ \forall p \in \check{Y}, x \in X_p\}$
- Opening: $\gamma_Y(X) = X \circ Y = (X \ominus Y) \oplus Y$
- Closing: $\varphi_Y(X) = X \bullet Y = (X \oplus Y) \ominus Y$

Opening is idempotent and anti-extensive. Closing is idempotent and extensive

Characterization of increasing operators on a complete lattice

Increasing operator Ψ in a lattice with partial order $\leq$:
$X, Y \in \mathcal{L}, X \leq Y \iff \Psi(X) \leq \Psi(Y)$

Theorem (Serra (1988), Heijmans & Ronse (1990))

Let us consider a complete lattice $\mathcal{L}$ and an increasing operator $\Psi : \mathcal{L} \to \mathcal{L}$, which preserves the greatest element $\top$; i.e., satisfies $\Psi(\top) = \top$.
Then Ψ is the supremum of a non-empty set of erosions $\mathcal{E}$:

$$\Psi = \bigvee_{\varepsilon \in \mathcal{E}} \varepsilon.$$

If the operator preserves the smallest element $\bot$; i.e., $\Psi(\bot) = \bot$, the operator Ψ can be written as an infimum of dilations

$$\Psi = \bigwedge_{\delta \in \mathcal{D}} \delta.$$

Matheron, Maragos and Banon–Barrera - MMBB

Theorem (Matheron, 1975)

Consider set operators on $\mathcal{P}(E)$. Let $\Psi : \mathcal{P}(E) \to \mathcal{P}(E)$ be a translation-equivariant (TE) increasing set operator. Then

$$\Psi(X) = \bigcup_{A \in Ker(\Psi)} X \ominus A = \bigcap_{B \in Ker(\Psi^*)} X \oplus \check{B}$$

where the kernel is $Ker(\Psi) = \{A \subseteq E \ : \ 0 \in \Psi(A)\}$

Theorem (Maragos, 1989)

Consider discrete set operators on $\mathcal{P}(\mathbb{Z}^n)$. Let $\Psi : \mathcal{P}(\mathbb{Z}^n) \to \mathcal{P}(\mathbb{Z}^n)$ be a TE, increasing and upper semi-continuous set operator. Then

$$\Psi(X) = \bigcup_{M \in Bas(\Psi)} X \ominus M = \bigcap_{N \in Bas(\Psi^*)} X \oplus \check{N}$$

where the basis (minimal kernel) of Ψ is

$$Bas(\Psi) = \{M \in Ker(\Psi) \ : \ [A \in Ker(\Psi) \text{ and } A \subseteq M] \implies A = M\}$$

Matheron, Maragos and Banon–Barrera - MMBB

Let us define a closed interval $[A, B] \in \mathcal{P}(\mathcal{E})$ by

$$[A, B] = \{X \subseteq E \ : \ A \subseteq X \subseteq B\}, \quad A, B \subseteq E$$

Theorem (Bannon–Barrera, 1991)

Any TE set operator (not necessarily increasing) $\Psi : \mathcal{P}(E) \to \mathcal{P}(E)$ can be represented as the union of sup-generating operators $\barwedge$ by pairs of sets that form closed intervals in its kernel:

$$\Psi(X) = \bigcup_{[A,B] \in \boldsymbol{Ker}(\Psi)} X \barwedge (A, B) = \bigcup_{[A,B] \in \boldsymbol{Ker}(\Psi)} [(X \ominus A) \cap (X^c \ominus B^c)]$$

Further, Ψ can be represented as the intersection of inf-generating operators $\veebar$ by pairs of reflected sets that form intervals in the kernel of its dual operator:

$$\Psi(X) = \bigcap_{[A,B] \in \boldsymbol{Ker}^*(\Psi)} X \veebar (\check{A}, \check{B}) = \bigcup_{[A,B] \in \boldsymbol{Ker}(\Psi)} [(X \oplus \check{A}) \cup (X^c \ominus (\check{B})^c)]$$

Matheron, Maragos and Banon–Barrera - MMBB

Maragos Theorem for function operators

Theorem (Maragos, 1989)

Consider an upper semi-continuous operator Ψ acting on an upper semi-continuous function f. Let $Bas(\Psi) = \{g_i\}_{i\in I}$ be its basis and $Bas(\bar{\Psi}) = \{h_j\}_{j\in J}$ the basis of the dual operator. If Ψ is a translation-equivariant and increasing operator then it can be represented as

$$\begin{aligned} \Psi(f)(x) &= \sup_{i\in I} (f \ominus g_i)(x) = \sup_{i\in I} \inf_{y\in\mathbb{R}^n} \{f(x+y) - g_i(y)\} \\ &= \inf_{j\in J} (f \oplus \check{h}_j)(x) = \inf_{j\in J} \sup_{y\in\mathbb{R}^n} \{f(x-y) + \check{h}_j(y)\} \end{aligned}$$

The converse is true. Given a collection of functions $\mathcal{B} = \{g_i\}_{i\in I}$ such that all elements of it are minimal in $(\mathcal{B}, \leq)$, the operator $\Psi(f) = \sup_{i\in I} \{f \ominus g_i\}$ is a translation-equivariant increasing operator whose basis is equal to $\mathcal{B}$.

1. Mathematical Morphology
2. A universal representation theorem for nonlinear (increasing) operators
3. Morphological activations
4. Group Morphology and equivariant universal representation
5. Convolution as an operator in an inf-semilattice
6. Morphological model of a CNN
7. Inf-semilattice of down/up-sampling and pyramids
8. Conclusions and Perspectives

Nonlinearities in Deep Learning as Dilations

- ReLU (Rectified Linear Unit) activation function:

$$\mathrm{ReLU}(f(x)) = \delta^{\mathrm{ReLU}}(f)(x) = \max(0, f(x))$$

A variant (Qiu et al. 2018)

$$\delta_{\alpha}^{\mathrm{ReLU}}(f)(x) = \max(0, f(x) + \alpha)$$

- Max-pooling (pooling window = stride):

$$\mathrm{MaxPooling}_{R \times R}(f)(x) = \delta_{R}^{\mathrm{MaxPool}}(f)(x) = \max_{y \in W_{R \times R}(x)} \{f(R \cdot x - y)\}$$

- Max-plus layer (morphological perceptron):

$$\delta_{b}^{MaxPlus}(\xi) = \max_{1 \le i \le d} \{\xi_i + b_i\}, \quad \xi, b \in \mathbb{R}^d$$

- Also Maxout and others...

Nonlinearities in Deep Learning as Dilations

- Why are dilations (with respect to the standard partial ordering $\leq$)? Because they satisfy
  - Increasing:

$$f \leq g \implies \delta^{\mathrm{ReLU}}(f) \leq \delta^{\mathrm{ReLU}}(g); \quad \delta_r^{\mathrm{MaxPool}}(f) \leq \delta_r^{\mathrm{MaxPool}}(g)$$

$$\xi \leq \eta \implies \delta_b^{MaxPlus}(\xi) \leq \delta_b^{MaxPlus}(\eta)$$

  - Commutation with supremum:

$$\delta^{\mathrm{ReLU}}(f \vee g) = \delta^{\mathrm{ReLU}}(f) \vee \delta^{\mathrm{ReLU}}(g); \delta_R^{\mathrm{MaxPool}}(f \vee g) = \delta_R^{\mathrm{MaxPool}}(f) \vee \delta_r^{\mathrm{MaxPool}}(g)$$

$$\delta_b^{MaxPlus}(\xi \vee \eta) = \delta_b^{MaxPlus}(\xi) \vee \delta_b^{MaxPlus}(\eta)$$

- The operators are also extensive, i.e., $f \leq \delta(f)$. ReLU is also idempotent, i.e., $\delta^{\mathrm{ReLU}}\delta^{\mathrm{ReLU}}(f) = \delta^{\mathrm{ReLU}}(f)$: δ^{ReLU} is a dilation and a closing

Morphological Activations (Velasco–Forero and A., 2022)

Proposition

Any increasing nonlinear activation function $\sigma : \mathbb{R} \to \mathbb{R}$ *in a neural network can be universally expressed as*

$$\sigma(z) = \min_{j \in J} \left[\max_{i \in I} \left\{ \beta_i^j z + \alpha_i^j \right\} \right] = \min_{j \in J} p_j(z)$$

where $p_j = \max_{i \in I} \left\{ \beta_i^j z + \alpha_i^j \right\}$ *is a PWL convex function.*

Proposition

Any increasing nonlinear pooling operator $\pi : \mathbb{R}^n \to \mathbb{R}^{n/R}$ *can be universally expressed as*

$$\pi(f)(x) = \min_{j \in J} \left[(f \oplus b_j) \right] (R \cdot x),$$

where $\{b_j\}_{j \in J}$ *is a family of structuring functions defining by transposition the basis of the dual operator to* π.

Morphological Activations (Velasco–Forero and A., 2022)

Tropical polynomial interpretation

- The max-affine function $p_j = \max_{i \in I} \left\{ \beta_i^j z + \alpha_i^j \right\}$ is a tropical polynomial such that in that geometry, the degree of the polynomial corresponds to the number of pieces of the PWL convex function
- The set of such polynomials constitutes the semiring $\mathbb{R}_{\max}$ of tropical polynomials
- Tropical geometry in the context of lattice theory and neural networks is an active area of research, however those previous works have not considered the use of minimal representation of tropical polynomials as generalized activation functions

Morphological Activations (Velasco–Forero and A., 2022)

Two alternative architectures of the MorphoActivation layer (Activation and Pooling Morphological Operator) $f \mapsto \Psi^{\mathrm{Morpho}} : \mathbb{R}^n \to \mathbb{R}^{(n/R)}$ either by composition $[\pi \circ \sigma(f)](x)$ or $[\sigma \circ \pi(f)](x)$ as follows:

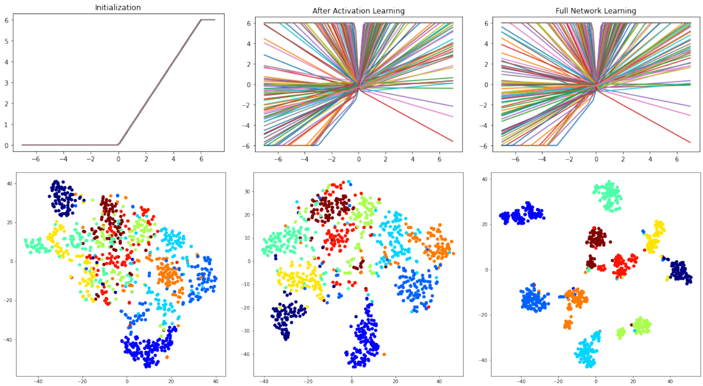
$$\Psi_1^{\mathrm{Morpho}}(f) = \min_{1 \le j \le M} \left\{ \delta_{R,b_j}^{\mathrm{MaxPool}} \left(\max_{1 \le i \le N} (\beta_i^j f + \alpha_i^j) \right) \right\},$$

$$\Psi_2^{\mathrm{Morpho}}(f) = \min_{1 \le i \le N} \left\{ \max_{1 \le j \le M} \left(\beta_i^j \delta_{R,b_i}^{\mathrm{MaxPool}}(f) + \alpha_i^j \right) \right\},$$

where

$$\begin{cases} \delta_{R,b_j}^{\mathrm{MaxPool}}(f)(x) = \delta_{b_j}(f)(R \cdot x), \quad \text{with} \\ \delta_{b_j}(f)(x) = (f \oplus b_j)(x) = \sup_{y \in W}\{f(x-y) + b_j(y)\} \end{cases}$$

Morphological Activations (Velasco–Forero and A., 2022)



Left: Random Initialization with (14%) of performance on the test set, We use a simplified version of proposed activation $\min(\max(\beta_0 x + \alpha_0, \beta_1 x + \alpha_1, \alpha_2), \alpha_3)$, with initialization $\max(\min(\mathrm{ReLU}(x), 6), -6)$ Center: Training only activations (92.38%), Right: Training Full Network (98,58%). Second Row: t-SNE visualization of last layer is the 10-classes MNIST prediction for a CNN.

Morphological Activations (Velasco–Forero and A., 2022)

	Fashion MNIST			CIFAR10			CIFAR100		
MaxPool(ReLU)	93.11			78.04			47.57		
MorphoActivation 1	N=2	N=3	N=4	N=2	N=3	N=4	N=2	N=3	N=4
M=2	-0.06	-0.05	-0.1	-0.42	0.02	-0.02	0.44	0.7	0.4
M=3	-0.14	-0.14	-0.06	-0.57	-0.4	-0.35	0.56	0.49	0.61
M=4	-0.02	-0.08	-0.01	0.05	-0.62	-0.5	0.41	0.35	0.73
MorphoActivation 2	N=2	N=3	N=4	N=2	N=3	N=4	N=2	N=3	N=4
M=2	0.04	-0.16	-0.12	1.84	2.02	1.49	3.31	3.5	3.45
M=3	0.08	-0.09	**0.12**	2.39	1.96	1.82	3.48	3.55	**3.86**
M=4	-0.02	0.09	-0.03	**2.49**	2.25	2.13	3.47	3.73	3.58

Relative difference with respect to our baseline (ReLU followed by a MaxPool). Architecture used is a CNN with two layers. ADAM optimizer with an early stopping with patience of ten iterations. Only Random Horizontal Flip has been used as image augmentation technique for CIFARs. The results are the average over three repetitions of the experiments.

1 Mathematical Morphology
2 A universal representation theorem for nonlinear (increasing) operators
3 Morphological activations
4 Group Morphology and equivariant universal representation
5 Convolution as an operator in an inf-semilattice
6 Morphological model of a CNN
7 Inf-semilattice of down/up-sampling and pyramids
8 Conclusions and Perspectives

Dilation and erosion on $\mathcal{P}(\mathbb{G})$

- Let us consider E is a homogeneous space under a group $\mathbb{G}$ acting transitively on E
- The object space of interest is the Boolean lattice $\mathcal{P}(E)$ of all subsets of E
- A mapping $\Psi : \mathcal{P}(\mathbb{G}) \to \mathcal{P}(\mathbb{G})$ is called $\mathbb{G}$-left-equivariant when, for all $g \in \mathbb{G}$, $\Psi(gG) = g\Psi(G)$, $\forall G \in \mathcal{P}(\mathbb{G})$. And similarly, a $\mathbb{G}$-right-equivariant implies for all $\forall G \in \mathcal{P}(\mathbb{G})$, $\Psi(Gg) = \Psi(G)g$
- Dilation and erosion on $\mathcal{P}(\mathbb{G})$ will be defined as the $\mathbb{G}$-equivariant mappings commuting with unions and intersections respectively
- Let H be a fixed subset of $\mathbb{G}$, called the group structuring element, we define the $\mathbb{G}$-left-equivariant dilation and erosion of G by H as

$$\delta^l_H(G) = G \oplus^l_{\mathbb{G}} H = \bigcup_{h \in H} Gh = \bigcup_{g \in G} gH = \{k \in \mathbb{G} : (k\check{H}) \cap G \neq \emptyset\},$$

$$\varepsilon^l_H(G) = G \ominus^l_{\mathbb{G}} H = \bigcap_{h \in H} Gh^{-1} = \{g \in \mathbb{G} : gH \subseteq G\},$$

where $gH = \{gh : h \in H\}$, $Hg = \{hg : h \in H\}$

Lifting and projections operators

- We work on the case of $\mathbb{G}$ is acting transitively on E Let the origin ω be an arbitrary point of E
- The lifting operator $\vartheta : \mathcal{P}(E) \to \mathcal{P}(\mathbb{G})$ is the mapping defined for any subset $X \in \mathcal{P}(E)$ as

$$\vartheta(X) = \{g \in \mathbb{G} \ : \ g\omega \in X\},$$

  associates to X all group elements which map the origin ω to an element of X
- The projection operator $\pi : \mathcal{P}(\mathbb{G}) \to \mathcal{P}(E)$ for any $G \in \mathcal{P}(\mathbb{G})$ as

$$\pi(G) = \{g\omega \ : \ g \in \mathbb{G}\},$$

  maps to each subset G of $\mathbb{G}$ the collection of points $g\omega \in E$, where g ranges over G
- The main benefit of creating these maps it that they translate the group action on X into multiplication in $\mathbb{G}$

$\mathbb{G}$-equivariant dilation and erosion on $\mathcal{P}(E)$

- Equivariant operator Ψ on $\mathcal{P}(E)$ by using the group operator $\tilde{\Psi}$:

$$\begin{array}{ccc} \mathcal{P}(\mathbb{G}) & \xrightarrow{\tilde{\Psi}} & \mathcal{P}(\mathbb{G}) \\ \uparrow \vartheta & & \downarrow \pi \\ \mathcal{P}(E) & \xrightarrow{\Psi} & \mathcal{P}(E) \end{array}$$

- Let us consider in particular the $\mathbb{G}$-equivariant dilation and erosion on $\mathcal{P}(E)$. For any set X and structuring element B, $X, B \in \mathcal{P}(E)$:

$$\delta_B^{\mathbb{G}}(X) = \pi\left[\vartheta(X) \oplus_{\mathbb{G}}^{l} \vartheta(B)\right] = \bigcup_{g \in \vartheta(X)} gB$$

$$\varepsilon_B^{\mathbb{G}}(X) = \pi_{\Sigma}\left[\vartheta(X) \ominus_{\mathbb{G}}^{l} \vartheta(B)\right] = \bigcap_{g \in \vartheta(X^c)} g\hat{B}^*$$

with $\hat{Y}^* = \left(\pi(\check{\vartheta}(Y))\right)^c$.

From group convolution to group dilations/erosions for functions (1/2) (Angulo, 2024)

- Group convolution (or more precisely, "correlation"): $\mathbb{G}$ a compact group, layer maps between $\mathbb{G}$-feature maps in $L_2(\mathbb{G})$ with kernel k

$$\left(f *_{\mathbb{G}} k\right)(g) = \int_{\mathbb{G}} f(h) k(g^{-1}h) dh$$

where dh is the left Haar measure on $\mathbb{G}$ and $f \in \mathcal{F}(\mathbb{G}, \mathbb{R})$

- Example in $SE(2)$:

$$\left(f *_{SE(2)} k\right)(x,\theta) = \int_{\mathbb{R}^2} \int_{\mathbb{S}^1} f(x',\theta') k\left(R_\theta^{-1}(x' - x), \theta' - \theta\right) dx' d\theta'$$

Note that the feature map f has been lifted to $\mathbb{G}$

From group convolution to group dilations/erosions for functions (2/2) (Angulo, 2024)

- Let us consider the counterpart group convolution in tropical semirings
- Set of functions, or feature maps $f : \mathbb{G} \to \bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty, +\infty\}$, with upper (or lower) semi-continuity on $\mathbb{G}$
- $\mathbb{G}$-equivariant max-plus dilation and erosion of function f by structuring function b, with $f, b \in \mathcal{F}(\mathbb{G}, \bar{\mathbb{R}})$,

$$\begin{aligned} (f \oplus_{\mathbb{G}} b)(g) &= \sup_{h \in \mathbb{G}} \left\{ f(h) + b(gh^{-1}) \right\} \\ (f \ominus_{\mathbb{G}} b)(g) &= \inf_{h \in \mathbb{G}} \left\{ f(h) - b(g^{-1}h) \right\} \end{aligned}$$

- $\mathbb{G}$-equivariant max-times-plus dilation and erosion of function f by pair of structuring functions $\{a, b\}$, $a(g) > 0$, $f, a, b \in \mathcal{F}(\mathbb{G}, \bar{\mathbb{R}})$,

$$\begin{aligned} (f \oplus_{\mathbb{G}} \{a, b\})(g) &= \sup_{h \in \mathbb{G}} \left\{ a(gh^{-1}) f(h) + b(gh^{-1}) \right\} \\ (f \ominus_{\mathbb{G}} \{a, b\})(g) &= \inf_{h \in \mathbb{G}} \left\{ \frac{1}{a(g^{-1}h)} \left(f(h) - b(g^{-1}h) \right) \right\} \end{aligned}$$

Representation of increasing $\mathbb{G}$-equivariant operators (1/2) (Angulo, 2024)

- Consider an increasing $\mathbb{G}$-equivariant group operator

$$\tilde{\Psi} : \mathcal{F}(\mathbb{G}, \bar{\mathbb{R}}) \to \mathcal{F}(\mathbb{G}, \bar{\mathbb{R}})$$

- The kernel of the operator $\tilde{\Psi}$ is given by:

$$\mathrm{Ker}(\tilde{\Psi}) = \left\{ b \ : \tilde{\Psi}(b)(\omega) \geq \perp \right\}, \ \ b \in \mathcal{F}(\mathbb{G}, \bar{\mathbb{R}})$$

and the corresponding morphological minimal basis of $\tilde{\Psi}$ is obtained from the kernel functions as its minimal elements with respect to the partial order $\leq$, i.e.,

$$\mathrm{Bas}(\tilde{\Psi}) = \left\{ b' \in \mathrm{Ker}(\tilde{\Psi}) \ : \ [b \in \mathrm{Ker}(\tilde{\Psi}) \text{ and } b \leq b'] \implies b = b' \right\}$$

This collection of functions can uniquely represent the $\tilde{\Psi}$ operator

Representation of increasing $\mathbb{G}$-equivariant operators (2/2) (Angulo, 2024)

Theorem

Consider a group operator $\tilde{\Psi}$ acting on an upper semi-continuous function f. Let $Bas(\tilde{\Psi}) = \{b_i\}_{i\in I}$ be its basis. If $\tilde{\Psi}$ is a $\mathbb{G}$-equivariant and increasing operator then it can be represented as

$$\tilde{\Psi}(f)(g) \quad = \quad \sup_{i\in I} \left[\left(f \ominus_{\mathbb{G}} b_i \right)(g) \right] = \sup_{i\in I} \inf_{h\in\mathbb{G}} \left\{ f(h) - b_i(g^{-1}h) \right\}$$

The converse is true.

Theorem

Let Ψ be an operator acting on upper semi-continuous functions $f : \mathcal{F}(E,\bar{\mathbb{R}}) \to \mathcal{F}(E,\bar{\mathbb{R}})$ on a homogeneous space E.
If Ψ is a $\mathbb{G}$-equivariant and increasing operator then it can be represented as

$$\Psi(f)(x) \quad = \quad \pi \left[\sup \left(\vartheta(f) \ominus_{\mathbb{G}} b_i \right) \right] = \pi \left[\sup \; \inf \left\{ \vartheta(f)(h) - b_i(g^{-1}h) \right\} \right]$$

1. Mathematical Morphology
2. A universal representation theorem for nonlinear (increasing) operators
3. Morphological activations
4. Group Morphology and equivariant universal representation
5. Convolution as an operator in an inf-semilattice
6. Morphological model of a CNN
7. Inf-semilattice of down/up-sampling and pyramids
8. Conclusions and Perspectives

Signal processing viewed as morphological operators (Keshet, 2000)[1] (1/2)

- The aim of a signal processing task is to address "simpler versions" of a given signal: the concept of simplicity varies with the approach but is generally associated to a partial ordering in the set of signals
- The partial ordering is usually related to some signal measure: spectral energy, entropy, information, SNR, smoothness, compressibility, etc.
- Either a least or greatest element exists, but not both: In many situations, one can find a signal that is the "simplest" one according to some given measure, e.g., the signal with least energy is the null function which also have the greatest compressibility. However, there is no single signal that corresponds to the greatest energy or the lowest compressibility

[1] R. Keshet. Traditional Signal Processing See From the Morphological Poset-Theoretical Point of View. Hewlett Packard Report, HPL-1999-39(R.1), 2000.

Signal processing viewed as morphological operators (Keshet, 2000)[2] (2/2)

- The corresponding partial ordering induces a complete semilattice structure to the signal poset: in general, it would be a inf-semilattice (simplification = is a loss of a magnitude)
- Erosion is naturally defined in the inf-semilattice: it is the increasing operator which commutes (distributes) over the infimum associated to the partial ordering
- Dilation, as a "generalized inverse", appears only as the adjoint operator to the erosion.Opening is by adjunction the compostion of the erosion followed by its adjoint dilation
- Operators based on erosion and opening, like residue by opening, semigroups and skeleton, are naturally defined too

[2] R. Keshet. Traditional Signal Processing See From the Morphological Poset-Theoretical Point of View. Hewlett Packard Report, HPL-1999-39(R.1), 2000.

Inf-semilattice of linear filtering (1/3)

- Convolution of a function $f \in \mathcal{L}^n = \mathrm{Fun}(\mathbb{R}^n, \mathbb{R})$ by a kernel $k \in \mathcal{L}^n$

$$\varepsilon_k^{\mathrm{Conv}}(f)(x) = (f * k)(x)$$

- Fourier transform of a function: $F(\omega) = \mathcal{F}(f)(\omega) = |F(\omega)| \exp(j\angle F(\omega))$, with modulus $|F(\omega)|$ and phase $\angle F(\omega)$ and *convolution theorem*

$$\varepsilon_k^{\mathrm{Conv}}(f)(x) = \mathcal{F}^{-1}\left\{F(\omega)K(\omega)\right\}(x)$$

Inf-semilattice of linear filtering (1/3)

- Convolution of a function $f \in \mathcal{L}^n = \mathrm{Fun}(\mathbb{R}^n, \mathbb{R})$ by a kernel $k \in \mathcal{L}^n$

$$\varepsilon_k^{\mathrm{Conv}}(f)(x) = (f * k)(x)$$

- Fourier transform of a function: $F(\omega) = \mathcal{F}(f)(\omega) = |F(\omega)| \exp\left(j\angle F(\omega)\right)$, with modulus $|F(\omega)|$ and phase $\angle F(\omega)$ and *convolution theorem*

$$\varepsilon_k^{\mathrm{Conv}}(f)(x) = \mathcal{F}^{-1}\left\{F(\omega)K(\omega)\right\}(x)$$

- Spectral partial ordering $\leq_{\mathcal{F}}$: For all pairs $f, g \in \mathcal{L}^n$

$$f \leq_{\mathcal{F}} g \iff \forall \omega \in \mathbb{R}^n, \begin{cases} |F(\omega)| \leq |G(\omega)|, & \text{and} \\ \angle F(\omega) = \angle G(\omega) \end{cases}$$

Inf-semilattice of linear filtering (1/3)

- Convolution of a function $f \in \mathcal{L}^n = \mathrm{Fun}(\mathbb{R}^n, \mathbb{R})$ by a kernel $k \in \mathcal{L}^n$

$$\varepsilon_k^{\mathrm{Conv}}(f)(x) = (f * k)(x)$$

- Fourier transform of a function: $F(\omega) = \mathcal{F}(f)(\omega) = |F(\omega)| \exp(j\angle F(\omega))$, with modulus $|F(\omega)|$ and phase $\angle F(\omega)$ and *convolution theorem*

$$\varepsilon_k^{\mathrm{Conv}}(f)(x) = \mathcal{F}^{-1}\left\{F(\omega)K(\omega)\right\}(x)$$

- Spectral partial ordering $\leq_{\mathcal{F}}$: For all pairs $f, g \in \mathcal{L}^n$

$$f \leq_{\mathcal{F}} g \iff \forall \omega \in \mathbb{R}^n, \begin{cases} |F(\omega)| \leq |G(\omega)|, & \text{and} \\ \angle F(\omega) = \angle G(\omega) \end{cases}$$

- Poset $(\mathcal{L}^n, \leq_{\mathcal{F}})$ is a complete inf-semilattice: The least element is the null function and the infimum $f \wedge_{\mathcal{F}} g$ is defined as

$$(f \wedge_{\mathcal{F}} g)(x) = \begin{cases} \mathcal{F}^{-1}\left\{\min\left(|F(\omega)|, |G(\omega)|\right)\exp\left(j\angle F(\omega)\right)\right\}, & \text{if } \angle F(\omega) = \angle G(\omega) \\ 0, & \text{otherwise} \end{cases}$$

Inf-semilattice of linear filtering (2/3)

Proposition (Keshet, 2000)

Linear convolution $\varepsilon_k^{Conv}(f)$ is an erosion in $(\mathcal{L}^n, \leq_{\mathcal{F}})$, i.e.,

- *(increasing)* $\forall f, g \in \mathcal{L}^n,\ f \leq_{\mathcal{F}} g \implies \varepsilon_k^{Conv}(f) \leq_{\mathcal{F}} \varepsilon_k^{Conv}(g)$
- *(commutation with infimum)* $\varepsilon_k^{Conv}(f \wedge_{\mathcal{F}} g) = \varepsilon_k^{Conv}(f) \wedge_{\mathcal{F}} \varepsilon_k^{Conv}(g)$

Proposition (Keshet, 2000)

The adjoint dilation of a linear convolution filtering $\delta_k^{*,Conv}$ in $(\mathcal{L}^n, \leq_{\mathcal{F}})$ is its generalized-inverse filtering operation, i.e.,

$$\delta_k^{*,Conv}(f) = \mathcal{F}^{-1}\left\{F(\omega)K^*(\omega)\right\}(x), \quad \text{where}$$

$$K^*(\omega) = \begin{cases} K(\omega)^{-1}, & \text{if } |K(\omega)| > 0 \\ 0, & \text{otherwise} \end{cases}$$

and by adjunction definition satisfies $\forall f \in \mathcal{L}^n$

$$\delta_k^{*,Conv}(f) = \bigwedge_{\mathcal{F}} \left\{g \in \mathcal{L}^n \ : \ f \leq_{\mathcal{F}} \varepsilon_k^{Conv}(g)\right\}$$

Inf-semilattice of linear filtering (3/3)

Proposition (Keshet, 2000)

The morphological opening associated to linear convolution filtering, $\gamma_k^{Conv}(f) = \delta_k^{*,Conv}\varepsilon_k^{Conv}(f)$, is an ideal linear filtering operator, i.e., idempotent $\gamma_k^{Conv}\gamma_k^{Conv}(f) = \gamma_k^{Conv}(f)$, anti-extensive $\gamma_k^{Conv}(f) \leq_{\mathcal{L}} f$

Inf-semilattice of linear filtering (3/3)

Proposition (Keshet, 2000)

The morphological opening associated to linear convolution filtering, $\gamma_k^{Conv}(f) = \delta_k^{*,Conv}\varepsilon_k^{Conv}(f)$, is an ideal linear filtering operator, i.e., idempotent $\gamma_k^{Conv}\gamma_k^{Conv}(f) = \gamma_k^{Conv}(f)$, anti-extensive $\gamma_k^{Conv}(f) \leq_{\mathcal{L}} f$

- The magnitude activation of the "feature map" associated to the kernel k follows an order structure based on the energy selection by k in the frequency domain
- If the spectral erosion in $(\mathcal{L}^n, \leq_{\mathcal{F}})$ is anti-extensive, i.e., $\forall f \in \mathcal{L}^n$

$$\varepsilon_k^{\mathrm{Conv}}(f) \leq_{\mathcal{F}} f$$

  the composition is ordered too, i.e.,

$$\varepsilon_{k_2}^{\mathrm{Conv}}\varepsilon_{k_1}^{\mathrm{Conv}}(f) \leq_{\mathcal{F}} \varepsilon_{k_1}^{\mathrm{Conv}}(f) \leq_{\mathcal{F}} f$$

  Effect of successive erosions: the signal becomes simpler and simpler by removing frequency information
- Obviously, semigroup property of kernels provides a semigroup of erosion. For instance, Gaussian convolution with kernel g_σ

Convolutions in DL (1/2)

- For the DL framework, a feature map input at a layer is typically a set of K scalar images f_i in $\mathcal{L}^2$. The convolution involves in general a 3D convolution: the K features maps are concatenated as a 3D image $f = (f_1, \cdots, f_K)$, $f \in \mathcal{L}^3$ and the kernel $k \in \mathcal{L}^3$. Then, the output 3D feature maps are contracted by linear combination to get a single 2D image as one of the new feature maps. The 3D convolution can be decomposed into separable $2D \times 1D$ convolutions
- For the sake of simplicity of the notation, without implying any change on the mathematical modelling, we consider here the case with a scalar image to the entry f which provided K maps (result of 2D convolutions), then combined using a neural network unit (perceptron without bias for now), i.e., f and k_i typically in $\mathcal{L}^2$

$$f \mapsto \sum_{i=1}^{K} w_i(f * k_i) = \sum_{i=1}^{K} w_i \varepsilon_{k_i}^{\mathrm{Conv}}(f) = \sum_{i=1}^{K} \varepsilon_{w_i k_i}^{\mathrm{Conv}}(f), \quad w_i \in \mathbb{R}$$

Convolutions in DL (2/2)

- This spectral operator in the inf-semilattice $(\mathcal{L}^n, \leq_{\mathcal{F}})$ is increasing:

$$\forall f, g \in \mathcal{L}^n,\;\; f \leq_{\mathcal{F}} g \implies \sum_{i=1}^{K} w_i \varepsilon_{k_i}^{\mathrm{Conv}}(f) \leq_{\mathcal{F}} \sum_{i=1}^{K} w_i \varepsilon_{k_i}^{\mathrm{Conv}}(g)$$

  and anti-extensive if the coefficients normalized to sum 1:

$$\forall f \in \mathcal{L}^n,\;\; \sum_{i=1}^{K} w_i \varepsilon_{k_i}^{\mathrm{Conv}}(f) \leq_{\mathcal{F}} f$$

- Supremum of spectral erosions : It should be notice the corresponding supremum in the spectral partial ordering $\bigvee_{\mathcal{F}}$ becomes standard summation only when the signal involved have disjoint Fourier spectra, i.e. $\cap_i 1_{|K_i(x)|>0} = \emptyset$. Therefore, in general

$$\sum_{i=1}^{K} w_i \varepsilon_{k_i}^{\mathrm{Conv}}(f) \neq \bigvee_{\mathcal{F}} \varepsilon_{w_i k_i}^{\mathrm{Conv}}(f)$$

  (taking the supremum will lead an universal approximation of operators in $(\mathcal{L}^n, \leq_{\mathcal{F}})$ using MMBB)

1 Mathematical Morphology

2 A universal representation theorem for nonlinear (increasing) operators

3 Morphological activations

4 Group Morphology and equivariant universal representation

5 Convolution as an operator in an inf-semilattice

6 Morphological model of a CNN

7 Inf-semilattice of down/up-sampling and pyramids

8 Conclusions and Perspectives

Morphological view of a Convolutional Neural Network (1/2)

- Typical one-layer architecture with K filters: a given feature map is obtained as

$$f \to \left(\sum_{i=1}^{K} w_i (f * k_i) + \alpha \right) \to \text{ReLU} \to \text{MaxPooling}_{R \times R}$$

Morphological view of a Convolutional Neural Network (1/2)

- Typical one-layer architecture with K filters: a given feature map is obtained as

$$f \rightarrow \left(\sum_{i=1}^{K} w_i (f * k_i) + \alpha \right) \rightarrow \text{ReLU} \rightarrow \text{MaxPooling}_{R \times R}$$

- Rewritten as a morphological composition of operators

$$\delta_R^{\text{MaxPool}} \delta^{\text{ReLU}} \left(\sum_{i=1}^{K} w_i \varepsilon_{k_i}^{\text{Conv}}(f) + \alpha \right) = \delta_R^{\text{MaxPool}} \delta_\alpha^{\text{ReLU}} \left(\sum_{i=1}^{K} w_i \varepsilon_{k_i}^{\text{Conv}}(f) \right)$$

Morphological view of a Convolutional Neural Network (1/2)

- Typical one-layer architecture with K filters: a given feature map is obtained as

$$f \to \left(\sum_{i=1}^{K} w_i (f * k_i) + \alpha \right) \to \mathrm{ReLU} \to \mathrm{MaxPooling}_{R\times R}$$

- Rewritten as a morphological composition of operators

$$\delta_R^{\mathrm{MaxPool}} \delta^{\mathrm{ReLU}} \left(\sum_{i=1}^{K} w_i \varepsilon_{k_i}^{\mathrm{Conv}}(f) + \alpha \right) = \delta_R^{\mathrm{MaxPool}} \delta_\alpha^{\mathrm{ReLU}} \left(\sum_{i=1}^{K} w_i \varepsilon_{k_i}^{\mathrm{Conv}}(f) \right)$$

- Alternative architecture using a max-plus layer

$$\delta_R^{\mathrm{MaxPool}} \delta^{\mathrm{ReLU}} \left(\max_{1\le i\le K} \{ \varepsilon_{k_i}^{\mathrm{Conv}}(f) + b_i \} \right) = \delta_R^{\mathrm{MaxPool}} \delta^{\mathrm{ReLU}} \delta_b^{\mathrm{MaxPlus}} \left(\{ \varepsilon_{k_i}^{\mathrm{Conv}}(f) \right.$$

Morphological view of a Convolutional Neural Network (2/2)

- The composition of dilations in the same complete lattice can often be factorized into a single operation

Morphological view of a Convolutional Neural Network (2/2)

- The composition of dilations in the same complete lattice can often be factorized into a single operation
- Activation and Pooling Dilation (APD):

$$\delta_{R;\alpha}^{\text{ActPool}}(f)(x) = \max_{y \in W_{R\times R}(x)} \{\max(0, f(R \cdot x - y) + \alpha)\}$$

Morphological view of a Convolutional Neural Network (2/2)

- The composition of dilations in the same complete lattice can often be factorized into a single operation
- Activation and Pooling Dilation (APD):

$$\delta_{R;\alpha}^{\text{ActPool}}(f)(x) = \max_{y \in W_{R\times R}(x)} \{\max(0, f(R\cdot x - y) + \alpha)\}$$

- A multilayer CNN will therefore be seen as the hierarchical composition (let's consider $r = 2$):

$$\begin{aligned} L_1 &= \delta_{2;\alpha_1}^{\text{ActPool}} \left(\sum_{i_1=1}^{K_1} w_{i_1} \varepsilon_{k_{i_1}}^{\text{Conv}}(f) \right) \\ L_n &= \delta_{2;\alpha_n}^{\text{ActPool}} \left(\sum_{i_n=1}^{K_n} w_{i_n} \varepsilon_{k_{i_n}}^{\text{Conv}}(L_{n-1}) \right) \\ &= \delta_{2;\alpha_n}^{\text{ActPool}} \left(\sum_{i_n=1}^{K_n} w_{i_n} \varepsilon_{k_{i_n}}^{\text{Conv}} \left[\delta_{2;\alpha_{n-1}}^{\text{ActPool}} \left(\sum_{i_{n-1}=1}^{K_{n-1}} w_{i_{n-1}} \varepsilon_{k_{i_{n-1}}}^{\text{Conv}}(L_{n-2}) \right) \right] \right) \end{aligned}$$

Fundamental fact: same partial ordering on two spaces (1/2)

- Let us consider two complex numbers z_1 and z_2 and let us introduce the following partial ordering in $\mathbb{C}$

$$z_1 \leq_{\mathbb{C}} z_2 \iff \begin{cases} |z_1| \leq |z_2|, & \text{and} \\ \angle z_1 = \angle z_2 \end{cases}$$

Same order than the spectral one in the Fourier transform domain

Fundamental fact: same partial ordering on two spaces (1/2)

- Let us consider two complex numbers z_1 and z_2 and let us introduce the following partial ordering in $\mathbb{C}$

$$z_1 \leq_{\mathbb{C}} z_2 \iff \begin{cases} |z_1| \leq |z_2|, & \text{and} \\ \angle z_1 = \angle z_2 \end{cases}$$

Same order than the spectral one in the Fourier transform domain

- Let us consider that for any $z_k = a_k + ib_k$ that the imaginary part is zero, $b_k = 0$, the previous partial ordering becomes

$$z_1 \leq_{\mathbb{C},\Im=0} z_2 \iff \begin{cases} |z_1| \leq |z_2|, & \text{and} \\ \text{sign}(z_1) = \text{sign}(z_2) \end{cases}$$

which is equivalent to median partial ordering $z_1 \preceq z_2$

Fundamental fact: same partial ordering on two spaces (2/2)

- Let us consider that for any $z_k = a_k + ib_k$ one has $b_k = 0$ and $a_k \geq 0$ thus $\mathrm{sign}(z_k) = +$, i.e., positive real value line $\mathbb{R}_+$, the complex partial ordering $z_1 \leq_{\mathbb{C},\Im=0,\Re\geq 0} z_2$ becomes the standard partial ordering $z_1 \leq z_2$
- The erosion ε_b in $(\mathcal{L}^n, \leq)$ is algebraically similar to the spectral erosion $\varepsilon_k^{\mathrm{Conv}}$ in $(\mathcal{L}^n, \leq_{\mathcal{F}})$

Fundamental fact: same partial ordering on two spaces (2/2)

- Let us consider that for any $z_k = a_k + ib_k$ one has $b_k = 0$ and $a_k \geq 0$ thus $\mathrm{sign}(z_k) = +$, i.e., positive real value line $\mathbb{R}_+$, the complex partial ordering $z_1 \leq_{\mathbb{C},\Im=0,\Re\geq 0} z_2$ becomes the standard partial ordering $z_1 \leq z_2$
- The erosion ε_b in $(\mathcal{L}^n, \leq)$ is algebraically similar to the spectral erosion $\varepsilon_k^{\mathrm{Conv}}$ in $(\mathcal{L}^n, \leq_{\mathcal{F}})$
- Fundamental fact. The same partial ordering $\leq_{\mathbb{C}}$ is always applied:
  - In the linear convolution operators, the poset structure is defined on the spectral space (i.e., $d-$dimensional complex-valued Fourier space)
  - In classical morphological operators or self-dual morphological operators, the poset structure is defined on the morphological space (i.e., $d-$dimensional (positive) real-valued intensity space)

Exploring advanced new nonlinear operators (1/2)

- ReLU and max-pooling can be replaced by a more general nonlinear operator defined by a morphological combination of activations, dilations and downsampling, using max-plus layer or its dual

Exploring advanced new nonlinear operators (1/2)

- ReLU and max-pooling can be replaced by a more general nonlinear operator defined by a morphological combination of activations, dilations and downsampling, using max-plus layer or its dual
- Morphological Activation of L structuring functions

$$f \mapsto \Psi^{\text{Morpho}}(f) = \min_{1 \leq j \leq J} \left\{ \delta_{R,b_j}^{\text{MaxPool}} \left(\max(0, \beta_j f) \right) + \alpha_j \right\}$$

where

$$\begin{cases} \delta_{R,b}^{\text{MaxPool}}(f)(x) = \delta_b(f)(R \cdot x), \quad \text{with} \\ \delta_b(f)(x) = (f \oplus b)(x) = \max_{y \in W}\{f(x-y) + b(y)\} \end{cases}$$

Exploring advanced new nonlinear operators (1/2)

- ReLU and max-pooling can be replaced by a more general nonlinear operator defined by a morphological combination of activations, dilations and downsampling, using max-plus layer or its dual
- Morphological Activation of L structuring functions

$$f \mapsto \Psi^{\text{Morpho}}(f) = \min_{1 \leq j \leq J} \left\{ \delta_{R,b_j}^{\text{MaxPool}} \left(\max(0, \beta_j f) \right) + \alpha_j \right\}$$

where

$$\begin{cases} \delta_{R,b}^{\text{MaxPool}}(f)(x) = \delta_b(f)(R \cdot x), \quad \text{with} \\ \delta_b(f)(x) = (f \oplus b)(x) = \max_{y \in W} \{ f(x-y) + b(y) \} \end{cases}$$

- The learnable structuring functions b_j play the same role as the kernels k_i in the convolutions. Note that one can have $R = 1$

Exploring advanced new nonlinear operators (2/2)

- Let us note the linear filtering part of the processing as

$$f \mapsto \Sigma^{\mathrm{Spectral}}(f) = \sum_{i=1}^{K} w_i \varepsilon_{k_i}^{\mathrm{Conv}}(f)$$

- A typical one-layer architecture with K filters in convolution and J structuring functions in APMO

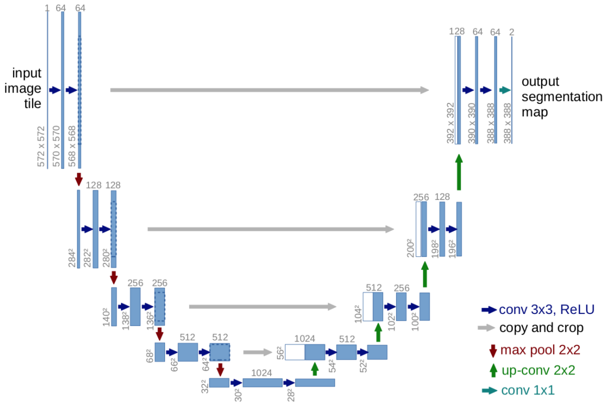
$$\begin{aligned} L_n &= \Psi_n^{\mathrm{Morpho}} \circ \Sigma_n^{\mathrm{Spectral}} (L_{n-1}) \\ &= \min_{1 \le j_n \le L_n} \left\{ \delta_{R_n, b_{j_n}}^{\mathrm{MaxPool}} \left[\max \left(0, \beta_{j_n} \sum_{i_n=1}^{K_n} w_{i_n} \varepsilon_{k_{i_n}}^{\mathrm{Conv}} (L_{n-1}) \right) \right] + \alpha_{j_n} \right\} \\ &= \Psi_n^{\mathrm{Morpho}} \circ \Sigma_n^{\mathrm{Spectral}} \circ \Psi_{n-1}^{\mathrm{Morpho}} \circ \Sigma_{n-1}^{\mathrm{Spectral}} (L_{n-2}) \end{aligned}$$

- The alternate processing between the feature detection by learning convolution kernels and the feature enhancement and selection (sparsity) by (learning) morphological parameters and structuring functions is probably the key in CNN...

1 Mathematical Morphology
2 A universal representation theorem for nonlinear (increasing) operators
3 Morphological activations
4 Group Morphology and equivariant universal representation
5 Convolution as an operator in an inf-semilattice
6 Morphological model of a CNN
7 Inf-semilattice of down/up-sampling and pyramids
8 Conclusions and Perspectives

More advanced DL architectures...

UNet (pyramidal fully convolutional network for segmentation)



But also ResNet (residual network)

Inf-semilattice of down/up-sampling and pyramids (1/5)

- Consider the set of discrete 2D real-valued images of $M \times N$ pixels, denoted by $\mathcal{L}_{M\times N}$
- Consider a dicrete version of the spectral partial ordering $\leq_{\mathcal{F}}$

$$f \leq_{\mathcal{F}} g \iff \forall i,j, \begin{cases} |F(i,j)| \leq |G(i,j)|, & \text{and} \\ \angle F(i,j) = \angle G(i,j) \end{cases}$$

where F and G are the discrete Fourier transforms (DFT) of f and g

- Given integer $R > 1$, the decimation operator of f by factor R is the mapping $\varepsilon_k^{\downarrow R,\mathrm{Conv}} : \mathcal{L}_{R\cdot M\times R\cdot N} \to \mathcal{L}_{M\times N}$ defined as

$$\varepsilon_k^{\downarrow R,\mathrm{Conv}}(f)(m,n) = \sum_{i=0}^{M-1}\sum_{j=0}^{N-1} f(R\cdot m - i, R\cdot n - j)k(i,j)$$

where $k(i,j)$ is the set of coefficients of the ideal π/R-cutoff low-pass filter

Inf-semilattice of down/up-sampling and pyramids (2/5)

Proposition (Keshet, 2000)

The decimation operator $\varepsilon_k^{\downarrow R,Conv}$ is an erosion between the complete inf-semilattices $(\mathcal{L}_{R\cdot M\times R\cdot N}, \leq_{\mathcal{F}})$ and $(\mathcal{L}_{M\times N}, \leq_{\mathcal{F}})$.
The adjoint dilation $\delta_k^{*\uparrow R,Conv} : (\mathcal{L}_{M\times N}, \leq_{\mathcal{F}}) \rightarrow (\mathcal{L}_{R\cdot M\times R\cdot N}, \leq_{\mathcal{F}})$ is the interpolation given by the up-sampling, followed by filtering with the same ideal π/R-cutoff low-pass filter k

Proposition (Keshet, 2000)

The corresponding morphological opening, i.e.,

$$\gamma_k^{\downarrow R\uparrow,Conv}(f) = \delta_k^{*\uparrow R,Conv}\varepsilon_k^{\downarrow R,Conv}(f),$$

is the linear filtering of f by k, i.e., $(f * k)(x)$ is the "best reconstruction" of f by (inverse) adjoint dilation after erosion.

Inf-semilattice of down/up-sampling and pyramids (3/5)

Morphological skeleton: Consider a series of erosions of increasing size $\{\varepsilon_i\}$, $\varepsilon_i(X) = X \ominus iB$ and $\varepsilon_i(X) = \varepsilon_1 \varepsilon_{i-1}(X)$. Given a set X, the skeleton $\mathrm{Skel}(X)$ is the union of the set of elements $S_i(X)$, seen as a decompostion/multiscale representation, defined as follows: $S_i(X) = \varepsilon_i(X) - \gamma_1 \varepsilon_i(X) = \varepsilon_i(X) - \delta_1 \varepsilon_{i+1}(X)$. Then, the reconstruction of the original set X from its skeleton is given by $X = \bigcup_i \delta_i\left(S_i\left(X\right)\right)$, also $\gamma_m(X) = \bigcup_{n \geq m} \delta_n\left(S_i\left(X\right)\right)$.

Proposition (Keshet, 2000)

The Laplacian pyramid $\{Lapl_i\}$ associated to f is a morphological skeleton in the complete inf-semilattices $(\mathcal{L}_{2^i \times 2^i}, \leq_{\mathcal{F}})$, $i = 0, \cdots, i_{\max}$. Formaly it is given by

$$\begin{aligned} Lapl_i(f) &= Gauss_i(f) - \delta_{g_\sigma}^{*\uparrow 2, Conv}\left(Gauss_{i+1}(f)\right) \\ &= Gauss_i(f) - \delta_{g_\sigma}^{*\uparrow 2, Conv} \varepsilon_{g_\sigma}^{\downarrow 2, Conv}\left(Gauss_i(f)\right) \\ &= \varepsilon_{g_\sigma}^{\downarrow 2, Conv} Gauss_{i-1}(f) - \gamma_{g_\sigma}^{\downarrow R\uparrow, Conv} \varepsilon_{g_\sigma}^{\downarrow 2, Conv} Gauss_{i-1}(f) \end{aligned}$$

where the Gaussian pyramid $\{Gauss_i(f)\}$ is given by $Gauss_i(f) = \varepsilon_{g_\sigma}^{\downarrow 2, Conv} Gauss_{i-1}(f)$.

Inf-semilattice of down/up-sampling and pyramids (4/5)

- In that case, we have in addition (identifying supremum $\vee_{\mathcal{F}}$ by standard sum):

$$\left[\mathrm{Gauss}_i(f) - \gamma_{g_\sigma}^{\downarrow R\uparrow,\mathrm{Conv}}(\mathrm{Gauss}_i(f))\right] \vee_{\mathcal{F}} \left[\gamma_{g_\sigma}^{\downarrow R\uparrow,\mathrm{Conv}}(\mathrm{Gauss}_i(f))\right] = \left[\mathrm{Gauss}_i(f) - \gamma_{g_\sigma}^{\downarrow R\uparrow,\mathrm{Conv}}(\mathrm{Gauss}_i(f))\right] + \left[\gamma_{g_\sigma}^{\downarrow R\uparrow,\mathrm{Conv}}(\mathrm{Gauss}_i(f))\right] = \mathrm{Gauss}_i(f)$$

- Let us consider the morphological residue by opening (aka top-hat):

$$\Gamma(f) = f - \gamma(f),$$

  note that $f \geq \gamma(f)$, therefore $\Gamma(f) \geq 0$ (potentially no activation is required associated to the corresponding partial ordering).
  For the skeleton, we can write $S_i(X) = \Gamma_1(\varepsilon_i(X))$

- In the Laplacian pyramid, one gets the residue by opening of the Gaussian pyramid:

$$\mathrm{Lapl}_i(f) = \Gamma_{g_\sigma}^{\downarrow R\uparrow,\mathrm{Conv}}(\mathrm{Gauss}_i(f))$$

Inf-semilattice of down/up-sampling and pyramids (5/5)

- In DL state-of-the-art, the decimation operator $\varepsilon_k^{\downarrow R,\mathrm{Conv}}$ is related to the (pooling) convolution with stride and the interpolation operator $\delta_k^{*\uparrow R,\mathrm{Conv}}$ with the (unpooling) dilated convolution
- In practice, one doesn't consider the ideal low-pass filter, i.e., in the spatial domain to the $sinc(x)$ function. The Gaussian kernel g_σ is a good approximation to an ideal low-pass filter:
  - But, $\gamma_{g_\sigma}^{\downarrow R\uparrow,\mathrm{Conv}}(f) \neq (f * g_\sigma)$
  - With the semigroup $(g_{\sigma_1} * g_{\sigma_2}) = g_{\sigma_1+\sigma_1}$, scale-space property in Gaussian pyramid, i.e., $\mathrm{Gauss}_i(f) = \varepsilon_{g_\sigma}^{\downarrow 2,\mathrm{Conv}}\mathrm{Gauss}_{i-1}(f)$, with $\mathrm{Gauss}_0(f) = f$
- In the general case where k is not a low-pass filter in the decimation operator $\varepsilon_k^{\downarrow R,\mathrm{Conv}}$, the interpolation adjoint operator $\delta_k^{*\uparrow R,\mathrm{Conv}}$ requires the generalized-inverse filtering in the convolution part (related to the deconvolution and transpose convolution in DL)

Morphological down/up-sampling and pyramids (1/5)

- Abstract pyramid structure:
  - Decomposition pyramid: From an initial signal f_0, approximations f_j of increasingly reduced size are computed by the analysis operator $\psi^{\downarrow}$
$$f_j = \psi^{\downarrow}(f_{j-1}), \quad j = 1, 2, \cdots, L$$
  - Residue and detail pyramid: An approximation error associated to f_{j+1} may be defined by taking the difference between f_j and an expanded version by the syntesis operator $\psi^{\uparrow}$
$$d_j = f_j - \psi^{\uparrow}(f_{j+1}), \quad j = 0, 2, \cdots, L$$
  - Iterative reconstruction: Original f_0 can be exactly reconstructed by the recursion
$$f_j = \psi^{\uparrow}(f_{j+1}) + d_j$$
  - Approximation: The reconstruction approximation to f_j is
$$\tilde{f}_j = \psi^{\uparrow}\left(\psi^{\downarrow}(f_j)\right)$$
- Pyramid condition: To guarantee that information lost during analysis can be recovered in the synthesis phase in a nonredundant way: $\psi^{\downarrow}\left(\psi^{\uparrow}(f)\right) = f$

Morphological down/up-sampling and pyramids (2/5)

- We now consider the general class of pyramids whose analysis/synthesis operator pairs have the form

$$\begin{aligned} \psi^{\downarrow}(f) &= \sigma^{\downarrow}(\eta(f)) \\ \psi^{\uparrow}(f) &= \xi(\sigma_c^{\uparrow}(f)) \end{aligned}$$

and $\sigma^{\downarrow}$ and $\sigma_c^{\uparrow}$ denote downsampling and upsampling by a factor of R in each spatial dimension, i.e.,

$$\begin{aligned} \sigma^{\downarrow}(f)(n) &= f(R \cdot n) \\ \sigma_c^{\uparrow}(f)(m) &= \begin{cases} f(n), & \text{if } m = R \cdot n \\ c, \text{otherwise} & \end{cases} \end{aligned}$$

- Adjoint pyramid by adjoint operators

Proposition (Goutsias and Heijmans, 2000)

If η and ξ are adjoint operators, the pair $(\psi^{\downarrow}, \psi^{\uparrow})$ forms an adjunction too

Morphological down/up-sampling and pyramids (3/5)

- "Increasing" equivariance to translation $\tau = \tau_y$ by pooling, where $(\tau_y f)(x) = f(x - y)$

Proposition (Goutsias and Heijmans, 2000)

If the pair of adjoint operators η and ξ are equivariant to translation, i.e., $\eta\tau = \tau\eta$ and $\xi\tau = \tau\xi$, one has for $R = 2$

$$\psi^{\downarrow}\tau^2 = \tau\psi^{\downarrow} \quad \psi^{\uparrow}\tau = \tau^2\psi^{\uparrow}$$

where $\tau^2 = \tau\tau$ denotes translation two times

Morphological down/up-sampling and pyramids (4/5)

- Goutsias–Heijmans morphological pyramid: A natural choice for the adjunction is $\eta = \varepsilon_b$ and $\xi = \delta_b^*$, so $\psi^{\downarrow R}$ is an erosion followed by decimation and $\psi^{\uparrow R}$ is an interpolation, with $c = \perp$ (minimum value) in $\sigma_c^{\uparrow R}$, followed by a dilation. The structuring function b is centred
  - Analysis (decimation) erosion by factor R

$$\begin{cases} \varepsilon_b^{\downarrow R}(f)(x) = \sigma^{\downarrow R}\left(\varepsilon_b(f)\right)(x), \quad \text{with} \\ \varepsilon_b(f)(x) = \min_{y\in W}\{f(y) - b(y-x)\} \end{cases}$$

  - Synthesis (interpolation) adjoint dilation by factor R

$$\begin{cases} \delta_b^{*\uparrow R}(f)(x) = \delta_b^*\left(\sigma_\perp^{\uparrow R}(f)\right)(x), \quad \text{with} \\ \varepsilon_b(f)(x) = \max_{y\in W}\{f(y) + b^*(y-x)\} \end{cases}$$

    where $b^*(x) = b(-x)$, and W is its support
  - Reconstruction is an opening: $\gamma_b^{\downarrow R\uparrow}(f) = \delta_b^*\sigma_\perp^{\uparrow R}\sigma^{\downarrow R}\varepsilon_b(f)$
- Sun–Maragos moprhological pyramid: analysis by $\eta = \delta_b^*\varepsilon_b = \gamma_b$ and synthesis by $\xi = \delta_b^*$

Morphological down/up-sampling and pyramids (5/5)

- Heijmans morphological pyramid: analysis by $\eta = \delta_b$ and synthesis by $\xi = \varepsilon_b^*$, with $c = \top$ (maximum value) in interpolation
  - Analysis (decimation) dilation by factor R

$$\begin{cases} \delta_b^{\downarrow R}(f)(x) = \sigma^{\downarrow R}\left(\delta_b(f)\right)(x), & \text{with} \\ \delta_b(f)(x) = \max_{y \in W}\{f(y) + b(x-y)\} \end{cases}$$

  - Synthesis (interpolation) adjoint erosion by factor R

$$\begin{cases} \varepsilon_b^{*\uparrow R}(f)(x) = \varepsilon_b^*\left(\sigma_\top^{\uparrow R}(f)\right)(x), & \text{with} \\ \varepsilon_b(f)(x) = \min_{y \in W}\{f(y) - b^*(y-x)\} \end{cases}$$

  - Reconstruction is a closing: $\varphi_b^{\downarrow R \uparrow}(f) = \varepsilon_b^* \sigma_\top^{\uparrow R} \sigma^{\downarrow R} \delta_b(f)$
- Max-pooling: $\delta_R^{\text{MaxPool}}(f) = \delta_b^{\downarrow R}(f)$, with a flat structuring function $b(z) = 0$ if $z \in W_{R \times R}$

Morphological view of UNet

Original architecture

$$
\begin{array}{ccc}
L_0^{\downarrow} = f & \longrightarrow & L_{n+1}^{\uparrow} = \delta^{\mathbf{ReLU}} \left(\sum_{i'_{n+1}=1}^{K'_{n+1}} w_{i'_{n+1}} \varepsilon^{\mathbf{Conv}}_{k_{i'_{n+1}}} \left(\left[\delta^{*\uparrow 2,\mathbf{Conv}}_{2\times 2}(L_n^{\uparrow}) \, , \, L_0^{\downarrow} \right] \right) \right) \\
\downarrow & & \uparrow \\
L_1^{\downarrow} = \delta^{\mathbf{ActPool}}_{2;\alpha_1} \left(\sum_{i_1=1}^{K_1} w_{i_1} \varepsilon^{\mathbf{Conv}}_{k_{i_1}}(f) \right) & \longrightarrow & L_n^{\uparrow} = \delta^{\mathbf{ReLU}} \left(\sum_{i'_n=1}^{K'_n} w_{i'_n} \varepsilon^{\mathbf{Conv}}_{k_{i'_n}} \left(\left[\delta^{*\uparrow 2,\mathbf{Conv}}_{2\times 2}(L_{n-1}^{\uparrow}) \, , \, L_1^{\downarrow} \right] \right) \right) \\
\downarrow & & \uparrow \\
L_2^{\downarrow} = \delta^{\mathbf{ActPool}}_{2;\alpha_2} \left(\sum_{i_2=1}^{K_2} w_{i_2} \varepsilon^{\mathbf{Conv}}_{k_{i_2}} \left(L_1^{\downarrow} \right) \right) & \longrightarrow & L_{n-1}^{\uparrow} = \delta^{\mathbf{ReLU}} \left(\sum_{i'_{n-1}=1}^{K'_{n-1}} w_{i'_{n-1}} \varepsilon^{\mathbf{Conv}}_{k_{i'_{n-1}}} \left(\left[\delta^{*\uparrow 2,\mathbf{Conv}}_{2\times 2}(L_{n-2}^{\uparrow}) \, , \, L_2^{\downarrow} \right] \right) \right) \\
\downarrow & & \uparrow \\
\cdots & \longrightarrow & \cdots \\
\downarrow & & \uparrow \\
L_n^{\downarrow} = \delta^{\mathbf{ActPool}}_{2;\alpha_n} \left(\sum_{i_n=1}^{K_n} w_{i_n} \varepsilon^{\mathbf{Conv}}_{k_{i_n}} \left(L_{n-1}^{\downarrow} \right) \right) & \longrightarrow & L_1^{\uparrow} = \delta^{\mathbf{ReLU}} \left(\sum_{i'_1=1}^{K'_1} w_{i'_1} \varepsilon^{\mathbf{Conv}}_{k_{i'_1}} \left(\left[\delta^{*\uparrow 2,\mathbf{Conv}}_{2\times 2}(L_0^{\uparrow}) \, , \, L_n^{\downarrow} \right] \right) \right) \\
\downarrow & & \uparrow
\end{array}
$$

$$
L_{n+1}^{\downarrow} = L_0^{\uparrow} = \delta^{\mathbf{ReLU}} \sum_{i_0=1}^{K_0} w_{i_0} \varepsilon^{\mathbf{Conv}}_{k_{i_0}} \left(L_n^{\downarrow} \right)
$$

Morphological view of UNet

More consistent architectures

- Linear down/up-sampling adjoint operators:

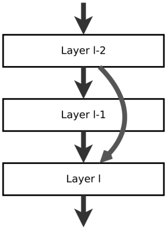
$$\begin{cases} L_n^{\downarrow} = \delta^{\mathrm{ReLU}}\, \varepsilon_{k_n}^{\downarrow 2,\mathrm{Conv}} \left(\sum_{i_n=1}^{K_n} w_{i_n} \varepsilon_{k_{i_n}}^{\mathrm{Conv}} \left(L_{n-1}^{\downarrow} \right) \right) \\ L_1^{\uparrow} = \delta^{\mathrm{ReLU}} \left(\sum_{i_1'=1}^{K_1'} w_{i_1'} \varepsilon_{k_{i_1'}}^{\mathrm{Conv}} \left(\left[\delta_{k_n}^{*\uparrow 2,\mathrm{Conv}}(L_0^{\uparrow}) \,,\, L_n^{\downarrow} \right] \right) \right) \end{cases}$$

- Morphological down/up-sampling adjoint operators:

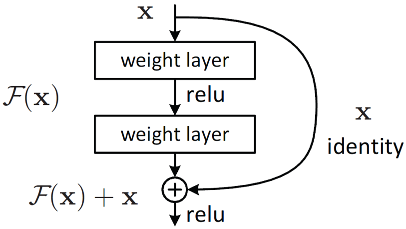
$$\begin{cases} L_n^{\downarrow} = \delta_{b_n}^{\downarrow 2}\, \delta^{\mathrm{ReLU}} \left(\sum_{i_n=1}^{K_n} w_{i_n} \varepsilon_{k_{i_n}}^{\mathrm{Conv}} \left(L_{n-1}^{\downarrow} \right) \right) \\ L_1^{\uparrow} = \delta^{\mathrm{ReLU}} \left(\sum_{i_1'=1}^{K_1'} w_{i_1'} \varepsilon_{k_{i_1'}}^{\mathrm{Conv}} \left(\left[\varepsilon_{b_n}^{*\uparrow 2}(L_0^{\uparrow}) \,,\, L_n^{\downarrow} \right] \right) \right) \end{cases}$$

Morphological view of ResNet

Canonical form of a residual neural network. A layer L_{n-1} is skipped over activation from L_{n-2}



Wikipedia



Doc Keras

$$L = \Psi_2^{\text{Morpho}} \left(\Sigma_2^{\text{Spectral}} \Psi_1^{\text{Morpho}} \, \Sigma_1^{\text{Spectral}} (f) + f \right)$$

if here we consider $\Sigma_2^{\text{Spectral}}$ is just a scalar $-w$, i.e.,

$$L = \Psi_2^{\text{Morpho}} \left(f - w \, \Psi_1^{\text{Morpho}} \, \Sigma_1^{\text{Spectral}} (f) \right)$$

Morphological view UNet and a proposal of UResNet

Let us note $\Sigma^{\text{Spectral}} = \sum_{i_n=1}^{K_n} w_{i_n} \varepsilon^{\text{Conv}}_{k_{i_n}}$, $\Psi^{\text{Morpho}} = \delta^{\text{ReLU}}$, $\Psi^{\text{Morpho}\downarrow} = \delta^{\downarrow 2}_{b_n} \delta^{\text{ReLU}}$ and $\Psi^{\text{Morpho}\uparrow} = \varepsilon^{*\uparrow 2}_{b_n}$

Variant of standard UNet

$$\begin{cases} L^{\downarrow}_n = \delta^{\downarrow 2}_{b_n} \Psi^{\text{Morpho}} \Sigma^{\text{Spectral}}_n \left(L^{\downarrow}_{n-1} \right) = \Psi^{\text{Morpho}\downarrow}_n \Sigma^{\text{Spectral}}_n \left(L^{\downarrow}_{n-1} \right) \\ L^{\uparrow}_1 = \Psi^{\text{Morpho}} \Sigma^{\text{Spectral}}_{1'} \left(\left[\varepsilon^{*\uparrow 2}_{b_n} (L^{\uparrow}_0), L^{\downarrow}_n \right] \right) = \Psi^{\text{Morpho}} \Sigma^{\text{Spectral}}_{1'} \left(\left[\Psi^{\text{Morpho}\uparrow}_{1'} (L^{\uparrow}_0), L^{\downarrow}_n \right] \right) \end{cases}$$

Residual UNet (UResNet)

$$\begin{cases} L^{\downarrow}_n = \Psi^{\text{Morpho}\downarrow}_n \Sigma^{\text{Spectral}}_n \left(L^{\downarrow}_{n-1} \right) \\ L^{\uparrow}_1 = \Psi^{\text{Morpho}} \left(L^{\downarrow}_n - \Sigma^{\text{Spectral}}_{1'} \left(\Psi^{\text{Morpho}\uparrow}_{1'} (L^{\uparrow}_0) \right) \right) \\ \quad = \Psi^{\text{Morpho}} \left(\Psi^{\text{Morpho}\downarrow}_n \Sigma^{\text{Spectral}}_n \left(L^{\downarrow}_{n-1} \right) - \Sigma^{\text{Spectral}}_{1'} \left(\Psi^{\text{Morpho}\uparrow}_{1'} (L^{\uparrow}_0) \right) \right) \end{cases}$$

1. Mathematical Morphology
2. A universal representation theorem for nonlinear (increasing) operators
3. Morphological activations
4. Group Morphology and equivariant universal representation
5. Convolution as an operator in an inf-semilattice
6. Morphological model of a CNN
7. Inf-semilattice of down/up-sampling and pyramids
8. Conclusions and Perspectives

Conclusions and Perspectives

- Integrated mathematical model of DL from mathematical morphology viewpoint: better algebraic theoretical understanding of operators helps to know
  - which ones can be "factorized" together
  - which are their "inverses"
  - which "composition rules" are algebraically consistent
  - which operators ones can be "switched" between linear and morphological, etc.
- We can study (and potentially) and improve some of the well-established DL architectures
- We can explore new architectures inspired from well-established morphological image processing pipelines
- We did not consider here the optimization viewpoint: which parameters are more effectively learned and in which optimization framework (optimizers, learning rate, etc.)?

A few open questions

How effective MMBB networks would be to learn the minimal basis of structuring functions approximating any nonlinear image transform? How the idea of hierarchical architectures from deep learning can be used in the case of MMBB networks?

Is there any information on order continuity, on invariance and fixed points, on decomposition and simplification, etc., which can be inferred from these unified algebraic models?

What is the expressiveness of deep spectral-morphological MMBB networks?