PDO-eConv: Partial Differential Operator Based Equivariant Convolutions

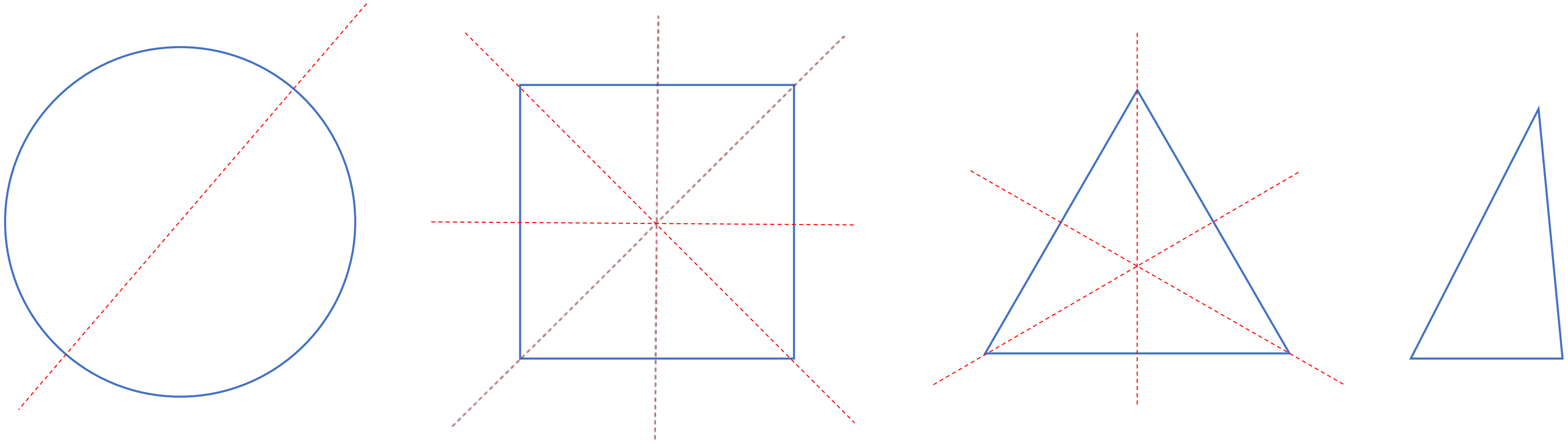
Zhouchen Lin
Peking University
July 28, 2020

Outline

1 | Background

Symmetries/Invariances are Everywhere

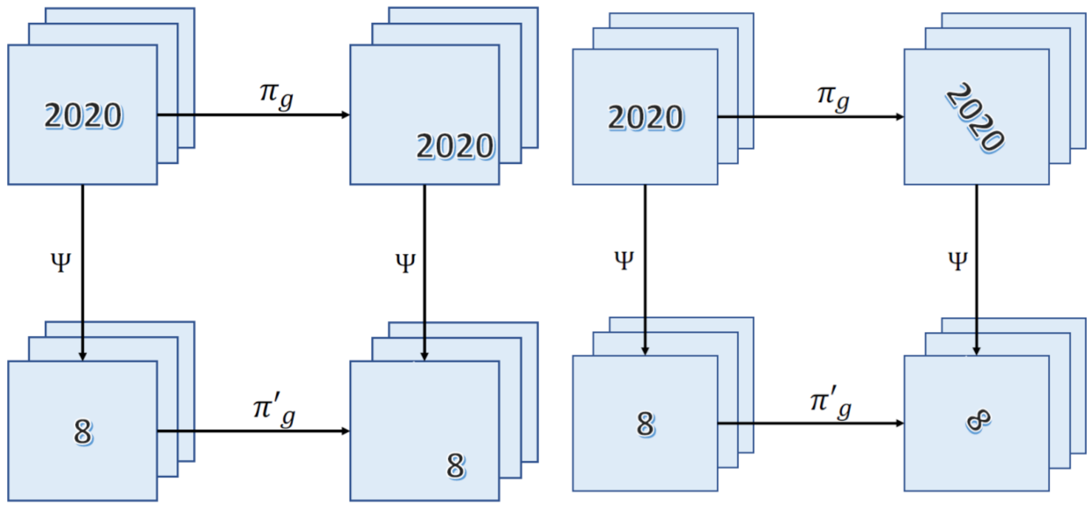
- What is symmetry or invariance?
  - A property that does not change under some transform



Group: O(2) | 4m | 3m | {Id}

Symmetries/Invariances are Common

- Physics
  - Newton’s Laws: Galilean Transformation
  - Maxwell’s Equations: Lorentzian Transformation
  - Special Relativity: FitzGerald–Lorentz–Einstein Transformation
  - General Relativity: Gauge Invariance
  - String Theory: Super-Symmetry
  - Higgs Particle: Local Gauge Invariance of Young-Mills Equation
- Visual Processing
  - Translational invariance
  - Rotational invariance
  - Perspective invariance



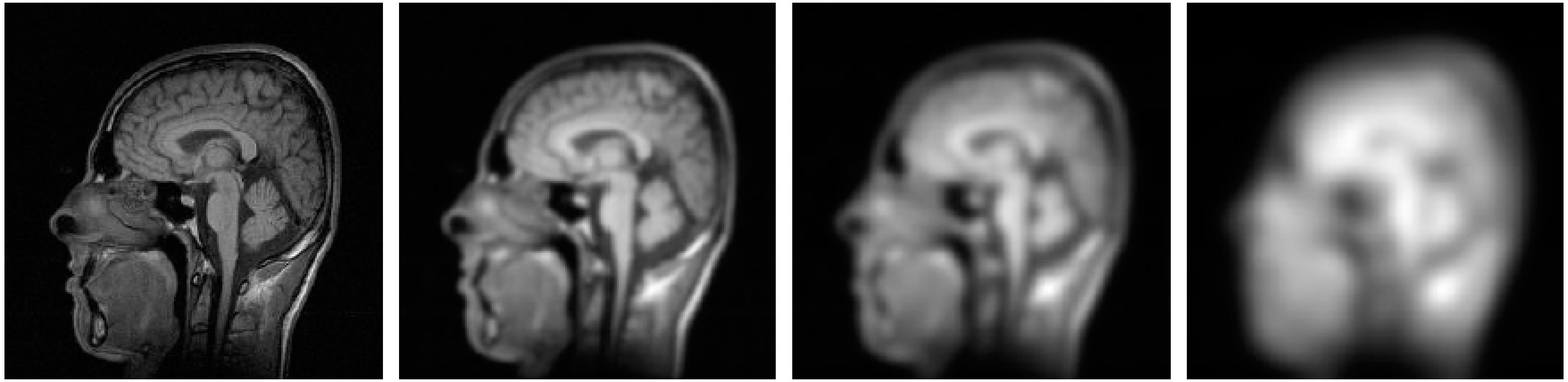
Also called **equivariance**

Symmetries/Invariances in PDEs

- PDEs in image processing

$$\begin{cases} \dfrac{\partial I}{\partial t} = \Delta I, \\ I|_{t=0} = I_0; \\ \left.\dfrac{\partial I}{\partial n}\right|_{\partial \Omega} = 0. \end{cases}$$

Input Image



A.K. Jain. *Partial Differential Equations and Finite-Difference Methods in Image Processing*, part 1. Journal of Optimization Theory and Applications, 23:65–91, 1977.

Symmetries/Invariances in PDEs

- Learning-based PDEs: **translational** and **rotational** invariances

$$\min J\left(\{O_m\}_{m=1}^{M}, \{a_j\}_{j=0}^{16}, \{b_j\}_{j=0}^{16}\right)$$
$$= \frac{1}{2}\sum_{m=1}^{M}\int_{\Omega}\left[O_m(x,1) - \tilde{O}_m(x)\right]^2 d\Omega + \frac{1}{2}\sum_{j=0}^{16}\lambda_j\int_0^1 a_j^2(t)dt + \frac{1}{2}\sum_{j=0}^{16}\mu_j\int_0^1 b_j^2(t)dt,$$

$$\begin{cases} O_t = \sum_{j=0}^{16} a_j(t)\,\mathrm{inv}_j(\rho, O), & (x,t) \in Q, \\ O = 0, & (x,t) \in \Gamma, \\ O|_{t=0} = O_0, & x \in \Omega, \end{cases}$$

$$\begin{cases} \rho_t = \sum_{j=0}^{16} b_j(t)\,\mathrm{inv}_j(O, \rho), & (x,t) \in Q, \\ \rho = 0, & (x,t) \in \Gamma, \\ \rho|_{t=0} = \rho_0, & x \in \Omega. \end{cases}$$

$(I_m, \tilde{O}_m)$ are training samples, where I_m is the input image and $\tilde{O}_m$ is the expected output image, $m = 1, 2, \cdots, M$.

Liu, Lin, Zhang, Tang, and Su, *Toward Designing Intelligent PDEs for Computer Vision: A Data-Based Optimal Control Approach*, Image and Vision Computing, 2013.

Symmetries/Invariances in PDEs

- Learning-based PDEs: **translational** and **rotational** invariances

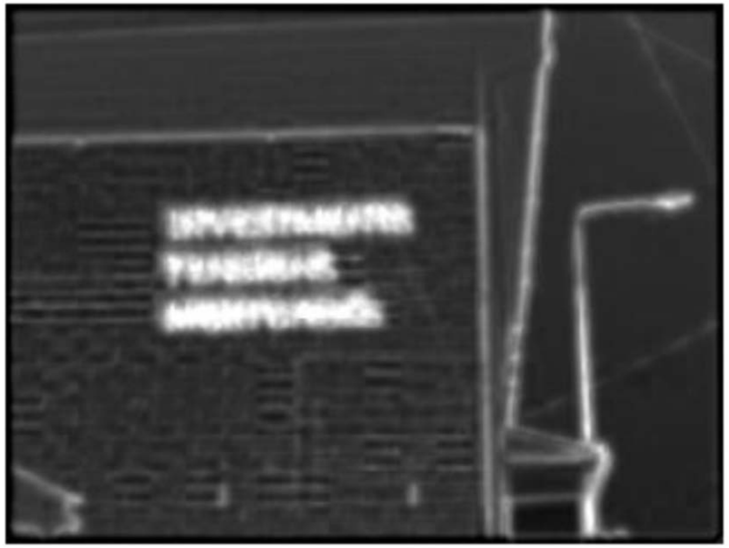
Table 1: Translationally and rotationally invariant fundamental differential invariants up to second order.

i	$\text{inv}_i(\rho, O)$
0,1,2	$1,\ \rho,\ O$
3,4,5	$\|\|\nabla\rho\|\|^2 = \rho_x^2 + \rho_y^2$, $(\nabla\rho)^t\nabla O = \rho_x O_x + \rho_y O_y$, $\|\|\nabla O\|\|^2 = O_x^2 + O_y^2$
6,7	$\text{tr}(\mathbf{H}_\rho) = \rho_{xx} + \rho_{yy}$, $\text{tr}(\mathbf{H}_O) = O_{xx} + O_{yy}$
8	$(\nabla\rho)^t\mathbf{H}_\rho\nabla\rho = \rho_x^2\rho_{xx}^2 + 2\rho_x\rho_y\rho_{xy}^2 + \rho_y^2\rho_{yy}^2$
9	$(\nabla\rho)^t\mathbf{H}_O\nabla\rho = \rho_x^2 O_{xx}^2 + 2\rho_x\rho_y O_{xy}^2 + \rho_y^2 O_{yy}^2$
10	$(\nabla\rho)^t\mathbf{H}_\rho\nabla O = \rho_x O_x \rho_{xx} + (\rho_y O_x + \rho_x O_y)\rho_{xy} + \rho_y O_y \rho_{yy}$
11	$(\nabla\rho)^t\mathbf{H}_O\nabla O = \rho_x O_x O_{xx} + (\rho_y O_x + \rho_x O_y)O_{xy} + \rho_y O_y O_{yy}$
12	$(\nabla O)^t\mathbf{H}_\rho\nabla O = O_x^2\rho_{xx} + 2O_x O_y \rho_{xy} + O_y^2\rho_{yy}$
13	$(\nabla O)^t\mathbf{H}_O\nabla O = O_x^2 O_{xx} + 2O_x O_y O_{xy} + O_y^2 O_{yy}$
14	$\text{tr}(\mathbf{H}_\rho^2) = \rho_{xx}^2 + 2\rho_{xy}^2 + \rho_{yy}^2$
15	$\text{tr}(\mathbf{H}_\rho\mathbf{H}_O) = \rho_{xx}O_{xx} + 2\rho_{xy}O_{xy} + \rho_{yy}O_{yy}$
16	$\text{tr}(\mathbf{H}_O^2) = O_{xx}^2 + 2O_{xy}^2 + O_{yy}^2$

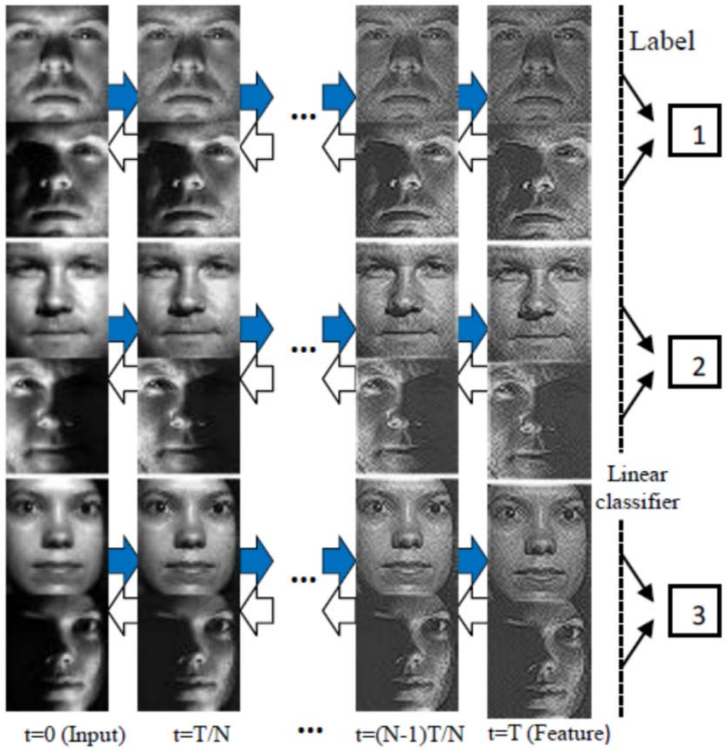
Liu, Lin, Zhang, Tang, and Su, *Toward Designing Intelligent PDEs for Computer Vision: A Data-Based Optimal Control Approach*, Image and Vision Computing, 2013.

Symmetries/Invariances in PDEs

- Learning-based PDEs: **translational** and **rotational** invariances
  - Applications in image processing, object detection, text detection, face recognition, …



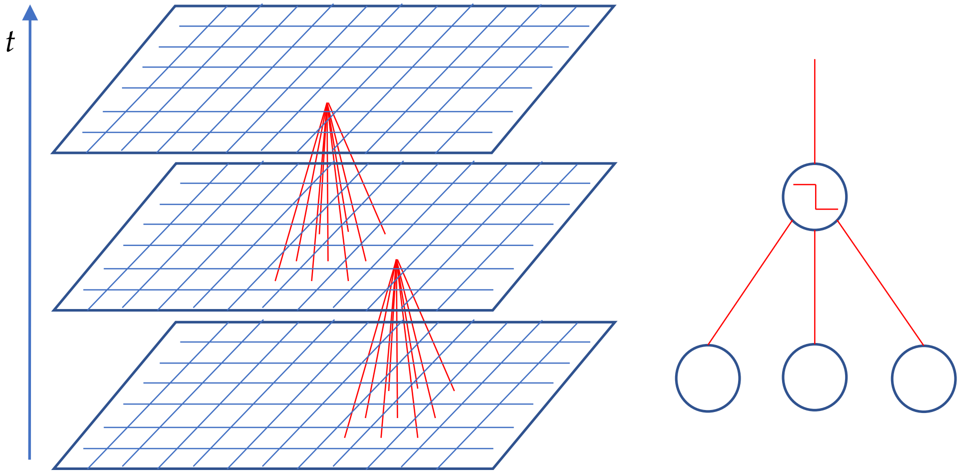
(f) Text confidence map



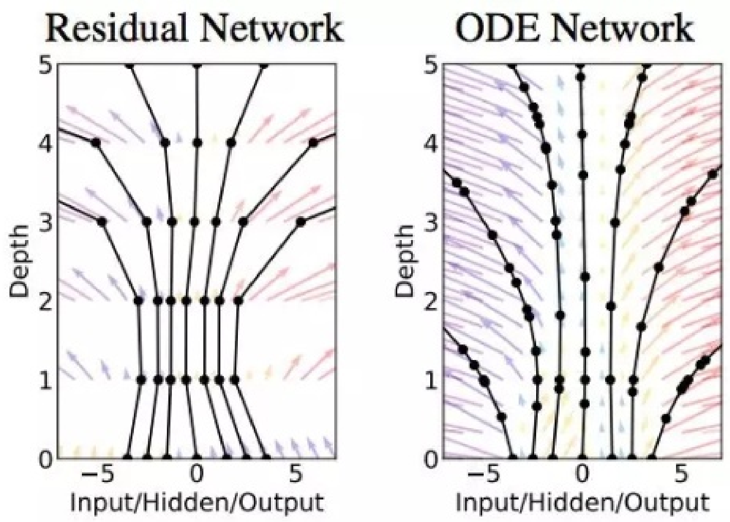
Liu, <u>Lin</u>, Zhang, Tang, and Su, *Toward Designing Intelligent PDEs for Computer Vision: A Data-Based Optimal Control Approach*, Image and Vision Computing, 2013.

Symmetries/Invariances in PDEs

- Learning-based PDEs: connections to DNNs



Vector Institute: Neural ODE



$$\mathbf{h}_{t+1} = \mathbf{h}_t + f(\mathbf{h}_t, \theta_t) \quad \Longleftrightarrow \quad \frac{\mathrm{d}\mathbf{h}(t)}{\mathrm{d}t} = f(\mathbf{h}_t, t, \theta(t))$$

ODE is insufficient as it only considers temporal relationship. A more complete theory should be based on PDE!

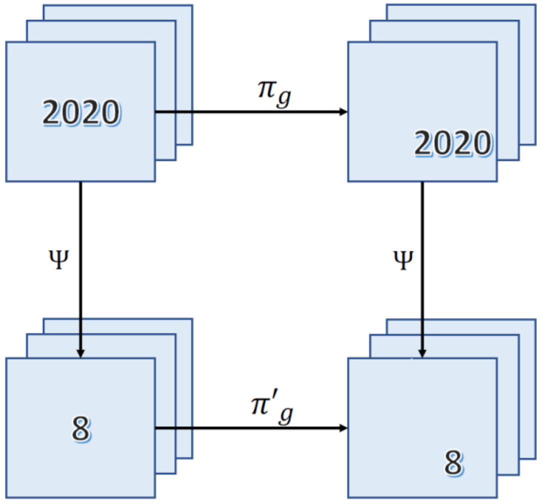
2 | Related Work

Group Equivariant CNNs

- Convolution is translational equivariant
- Group equivariance:

$$\Psi[\pi_g[f]] = \pi'_g[\Psi[f]],$$

where $g \in G$ and G is a translation group.



But convolution is **not rotational equivariant**!

Use a larger group G that contains both translation and rotation!

Cohen and Welling. *Group Equivariant Convolutional Networks*. ICML, 2016.

Group Equivariant CNNs

- Using correlation instead of convolution

$$[f * \psi^i](x) = \sum_{y \in \mathbb{Z}^2} \sum_{k=1}^{K'} f_k(y) \psi_k^i(x - y).$$

⇒

$$[f \star \psi^i](x) = \sum_{y \in \mathbb{Z}^2} \sum_{k=1}^{K'} f_k(y) \psi_k^i(y - x).$$

If one employs convolution in the forward pass, the correlation will appear in the backward pass when computing gradients, and vice versa.

G-Equivariant correlation:

- The first layer of a G-CNN:

$$[f \star \psi](g) = \sum_{y \in \mathbb{Z}^2} \sum_{k} f_k(y) \psi_k\left(g^{-1} y\right)$$

A feature map is a function on group.

- For all layers after the first:

$$[f \star \psi](g) = \sum_{h \in G} \sum_{k} f_k(h) \psi_k\left(g^{-1} h\right)$$

- Equivariant on group G:

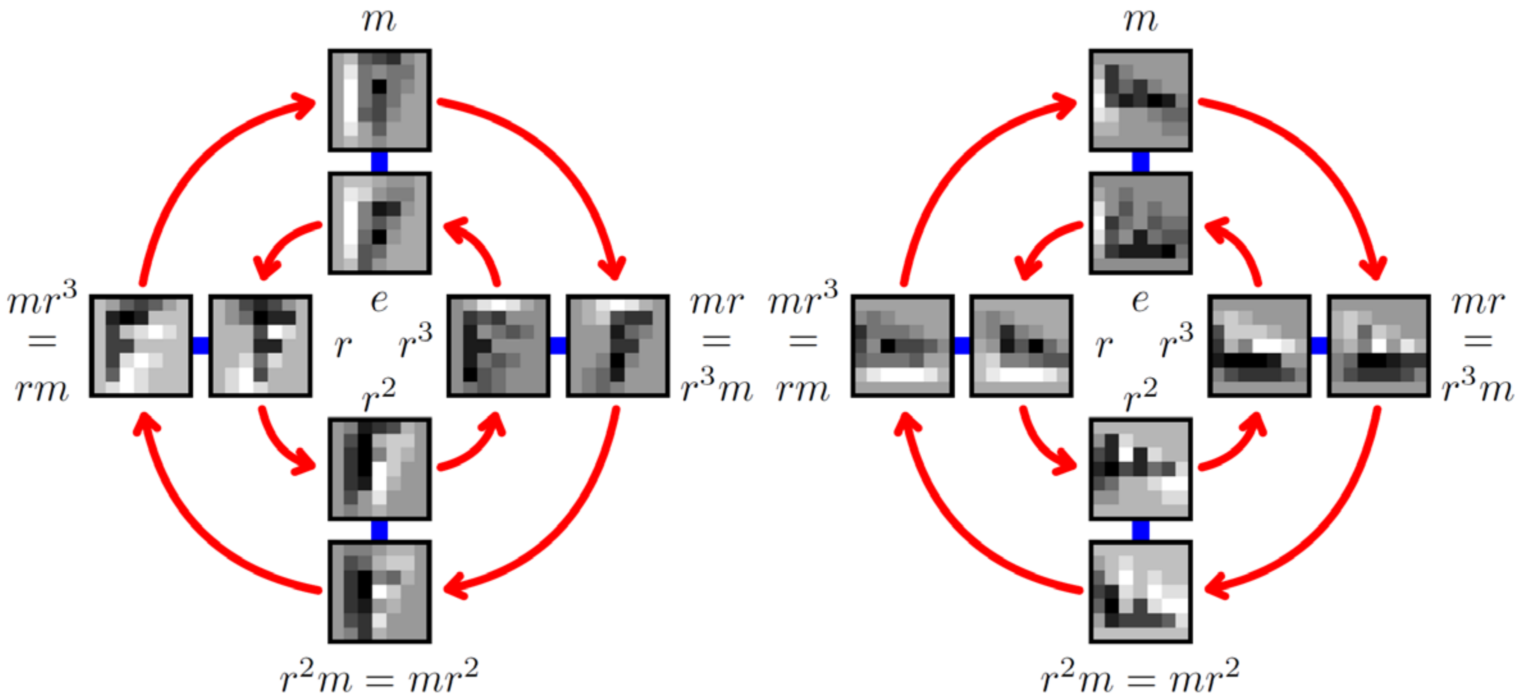
$$[[L_u f] \star \psi](g) = [L_u [f \star \psi]](g)$$

$$[L_u f](g) = f(u^{-1} g)$$

Cohen and Welling. *Group Equivariant Convolutional Networks*. ICML, 2016.

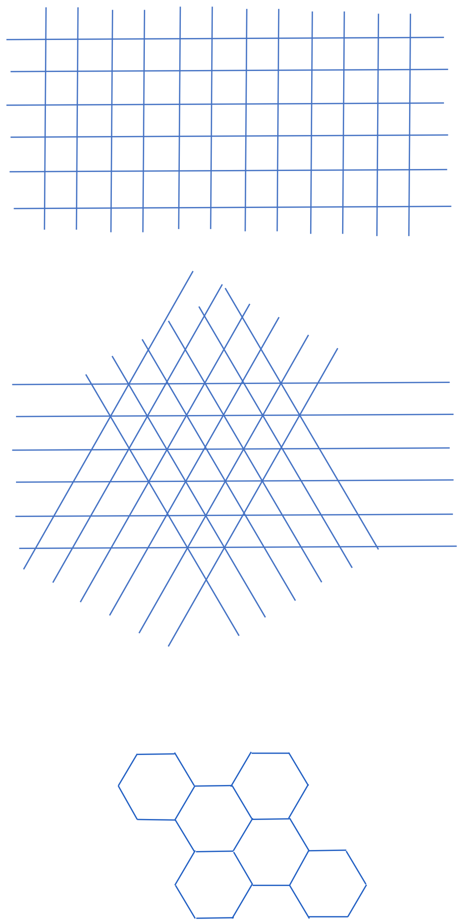
Group Equivariant CNNs

- The feature map defined on $G = p4m$



can only deal with a 4-fold rotational symmetry

2D lattices with symmetry



Cohen and Welling. *Group Equivariant Convolutional Networks*. ICML, 2016.

Follow-up Works

- Exploit more symmetries
  - Zhou et al. utilized bilinear interpolation to help produce feature maps at different orientations.
  - Weiler et al. employed harmonics as steerable filters to achieve exact equivariance w.r.t. larger groups in the continuous domain.

However, these methods cannot preserve strong equivariance in the discrete domain.

- In our work,
  - We use partial differential operators (PDOs) to design a system equivariant to $\boldsymbol{E(n)}$.
  - We use numerical schemes to discretize this system and obtain PDO-eConvs, which achieve a quadratic order equivariance approximation in the discrete domain.

Zhou et al. *Oriented Response Networks*. CVPR, 2017.
Weiler et al. *Learning Steerable Filters for Rotation Equivariant CNNs*. CVPR, 2018.

Mathematical Framework

- The Euclidean group $E(n) = \mathbb{R}^n \rtimes O(n)$. We use (x, A) to represent the element in $E(n)$, where x and A represent a translation and a orthogonal transformation, respectively.

- Actions on functions:

  - For inputs r defined on $\mathbb{R}^n$ and $\widetilde{A} \in O(n)$:

$$\forall x \in \mathbb{R}^n, \quad \pi^R_{\widetilde{A}}[r](x) = r(\widetilde{A}^{-1}x).$$

  - For intermediate feature maps e defined on $E(n)$:

$$\forall a \in E(n), \quad \pi^E_{\widetilde{A}}[e](a) = e(\widetilde{A}^{-1}a),$$

i.e.,

$$\pi^E_{\widetilde{A}}[e](x, A) = e(\widetilde{A}^{-1}x, \widetilde{A}^{-1}A).$$

Shen et al., *PDO-eConv: Partial Differential Operator Based Equivariant Convolutions*, ICML 2020.

Group Equivariant Differential Operators

- $H(u_1, u_2, \cdots, u_n; \boldsymbol{\beta})$ is a polynomial of n variables parameterized by $\boldsymbol{\beta}$. As a polynomial of PDOs, $H(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \ldots, \frac{\partial}{\partial x_n}; \boldsymbol{\beta})$ is a linear combination of PDOs parameterized by $\boldsymbol{\beta}$.

e.g., if $H(u) = u^2$, then $H(\frac{\partial}{\partial x}) = \frac{\partial^2}{\partial x^2}$

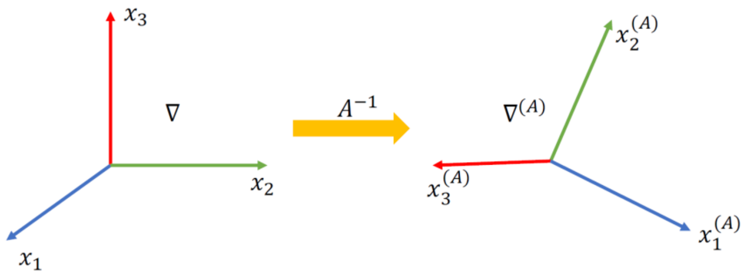
- We use orthogonal matrices to transform these PDOs

$$\chi^{(A)} = H\left(\frac{\partial}{\partial x_1^{(A)}}, \frac{\partial}{\partial x_2^{(A)}} \ldots, \frac{\partial}{\partial x_n^{(A)}}; \boldsymbol{\beta}\right),$$

where

$$\nabla^{(A)} = A^{-1}\nabla$$

$$\left(\frac{\partial}{\partial x_1^{(A)}}, \frac{\partial}{\partial x_2^{(A)}}, \cdots, \frac{\partial}{\partial x_n^{(A)}}\right)^T = A^{-1}\left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \cdots, \frac{\partial}{\partial x_n}\right)^T$$



and A is an orthogonal matrix.

Shen et al., *PDO-eConv: Partial Differential Operator Based Equivariant Convolutions*, ICML 2020.

Two Key Differential Operators

- we use Ψ to deal with inputs, which maps an input r defined on $\mathbb{R}^n$ to a feature map defined on $E(n)$:

$$\forall (x, A) \in E(n), \quad \Psi[r](x, A) = \chi^{(A)}[r](x).$$

- we use Φ to deal with the intermediate feature maps, which maps one feature map e on $E(n)$ to another feature map defined on $E(n)$:

$$\forall (x, A) \in E(n), \quad \Phi[e](x, A) = \int_{O(n)} \chi_B^{(A)} \, [e](x, AB) d\nu(B),$$

- Equivariance:

$$\Psi\left[\pi_{\widetilde{A}}^{R}[r]\right] = \pi_{\widetilde{A}}^{E}\left[\Psi[r]\right], \quad \Phi\left[\pi_{\widetilde{A}}^{E}[e]\right] = \pi_{\widetilde{A}}^{E}\left[\Phi[e]\right].$$

It can be easily extended to subgroups $G \subseteq E(n)$.

Shen et al., *PDO-eConv: Partial Differential Operator Based Equivariant Convolutions*, ICML 2020.

Discretization

- Choice of H:

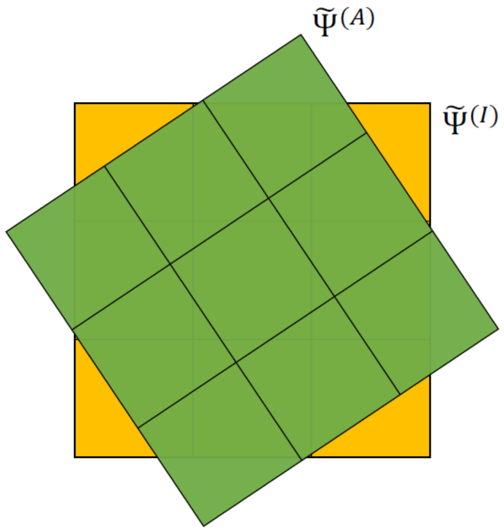
  We can translate a 3×3 convolutional filter to a differential operator

$$\mathcal{D} = \beta_1 + \beta_2\partial_x + \beta_3\partial_y + \beta_4\partial_{xx} + \beta_5\partial_{xy} + \beta_6\partial_{yy} + \beta_7\partial_{xxy} + \beta_8\partial_{xyy} + \beta_9\partial_{xxyy}.$$

  So we choose

$$H(u, v; \boldsymbol{\beta}) = \beta_1 + \beta_2 u + \beta_3 v + \beta_4 u^2 + \beta_5 uv + \beta_6 v^2 + \beta_7 u^2 v + \beta_8 uv^2 + \beta_9 u^2 v^2.$$

- Choice of group G: pn or pnm on the 2D plane.
- We use numerical schemes of PDOs to discretize the differential operators.
- Usage of PDO-eConvs: replace conventional convolutions by PDO-eConvs.



5x5 filters are sufficient! **The approximation error is of the second order!**

Shen et al., *PDO-eConv: Partial Differential Operator Based Equivariant Convolutions*, ICML 2020.

Experiments

- MNIST-rot is the most commonly used dataset for validating rotation-equivariant algorithms. It contains the handwritten digits of the classical MNIST, rotated by a random angle from 0 to 2π (full angle).

Method	Test Error	Method	Test Error
H-Net	1.69	OR-TIPooling	1.54
RotEqNet	1.09	PTN-CNN	0.89
E2CNN	0.716	SFCNN	0.714
PDO-eConv (ours)	**0.709**		

Comparable result with SFCNN using only 10% parameters (0.65M vs. 6.5M) and much less computational cost (5x5 vs. 7x7 & 9x9).

Shen et al., *PDO-eConv: Partial Differential Operator Based Equivariant Convolutions*, ICML 2020.

Experiments

- Natural image classification: CIFAR-10/100

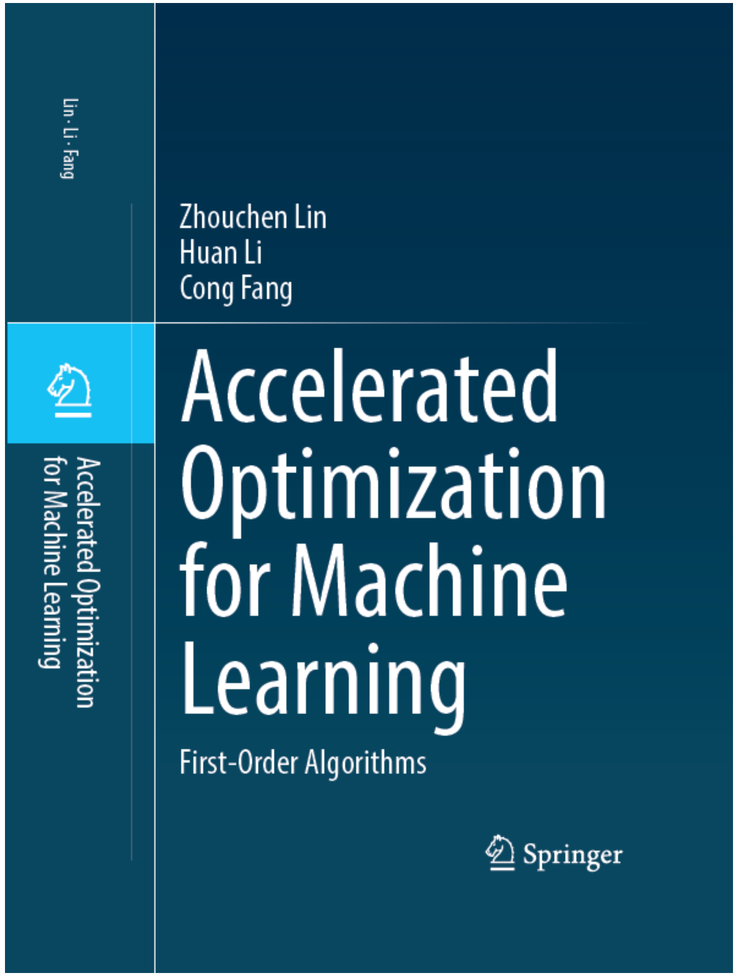
Method	G	Depth	C10	C100	params
ResNet	$\mathbb{Z}^2$	26	11.5	31.66	0.37M
HexaConv	$p6$	26	9.98	-	0.34M
	$p6m$	26	8.64	-	0.34M
PDO-eConv (ours)	$p6$	26	5.65	27.13	0.36M
	$p6m$	26	5.38	27.00	0.37M
ResNet	$\mathbb{Z}^2$	44	5.61	24.08	2.64M
G-CNN	$p4m$	44	4.94	23.19	2.62M
PDO-eConv (ours)	$p8$	44	3.68	20.01	2.62M
ResNet	$\mathbb{Z}^2$	1001	4.92	22.71	10.3M
Wide ResNet	$\mathbb{Z}^2$	26	4.00	19.25	36.5M
G-CNN	$p4m$	26	4.17	-	7.2M
PDO-eConv (ours)	$p8$	26	**3.50**	**18.40**	4.6M

PDO-eConvs can be viewed as introducing a weight sharing scheme across channels in Wide ResNet, which can not only save parameters, but also improve the performance remarkably.

Shen et al., *PDO-eConv: Partial Differential Operator Based Equivariant Convolutions*, ICML 2020.

Thank you!

- My webpage: https://zhouchenlin.github.io
- ZERO-lab webpage: https://zero-lab-pku.github.io



Recruitments

PKU: PostDoc (**270K** RMB/year) and Faculty

Samsung Beijing AI Lab: Researcher

之江Lab: Researcher and PostDoc

All in **machine learning** related areas!