3D Steerable CNNs: Learning Rotationally Equaivariant Features in Volumetric Data

Credit: Maurice Weiler[1], Mario Geiger[2], Max Welling[1], Wouter Boomsma[3], Taco Cohen[4]

[1]University of Amsterdam

[2]EPFL

[3]University of Copenhagen

[4]Qualcomn AI Research

Presenter: Fuwen Tan
https://qdata.github.io/deep2Read

Rotationally Equivariant Features

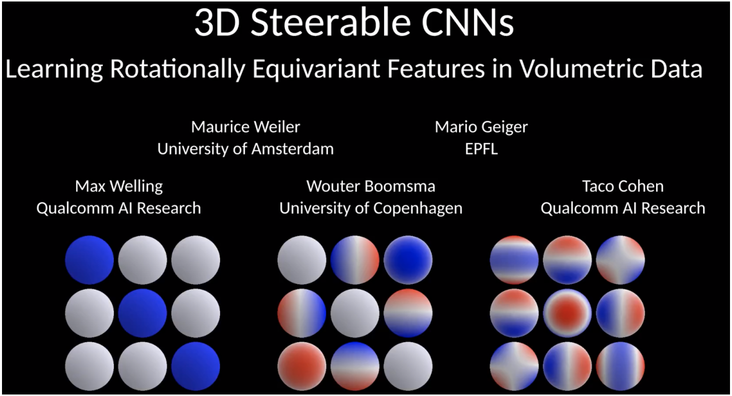

Figure: `https://www.youtube.com/watch?v=ENLJACPHSEA`

Benefits

- Data efficiency
- Therefore, less parameters

Definition

$f : \mathbb{R}^3 \to \mathbb{R}^{K_n}$	3D feature map of the n-th layer
$g = tr \in SE(3)$	3D rigid transformation on $\mathbb{R}^3$

Rotation of the feature map f

$$[\pi(r)f](x) \quad := \quad \rho(r)f(r^{-1}x)$$

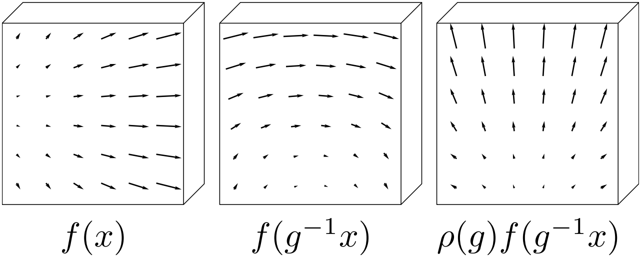

Figure: To transform a vector field (L) by a 90° rotation g, first move each arrow to its new position (C), keeping its orientation the same, then rotate the vector itself (R). This is described by the induced representation $\pi = \mathrm{Ind}_{\mathrm{SO}(3)}^{\mathrm{SE}(2)} \rho$, where $\rho(g)$ is a 3×3 rotation matrix that mixes the three coordinate channels.

Transformations of the feature map f

$$
\begin{aligned}
[\pi(tr)f](x) &:= \rho(r) f(r^{-1}(x - t)) \\
\rho(r) &: \mathbb{R}^K \to \mathbb{R}^K, \text{ invertible}
\end{aligned}
$$

Formulation

A filter κ (e.g. a $(3 \times 3 \times K_n \times K_{n+1})$ filter) is SE(3) equivariant if

$$\kappa \cdot [\pi_1(g)f] = \pi_2(g)[\kappa \cdot f]$$

- They prove that the space of κ is a subspace of 3D convolutional filter
- They prove the space of κ is linear, and can be represented as a linear combination of a set of basic filters

Representation of a group

- ρ is an invertible $n \times n$ matrix parameterized by a group element (e.g. rotation r).
- For ρ to be called a representation of G, it has to satisfy $\rho(gg') = \rho(g)\rho(g')$, where gg' denotes the composition of two transformations $g, g' \in G$, and $\rho(g)\rho(g')$ denotes matrix multiplication.

A concrete example

- 3×3 matrix A
- Transformation: $A \mapsto R(r)AR(r)^T$, $R(r)$: 3×3 rotation
- Kronecker / tensor product:
  $\mathrm{vec}(A) \mapsto [R(r) \otimes R(r)]\, \mathrm{vec}(A) \equiv \rho(r)\, \mathrm{vec}(A)$.
- $\rho(r)$ is a 9-dimensional representation of $\mathrm{SO}(3)$

Decomposition of $\rho(r)$

$$\rho(r) = Q^{-1}\left[\bigoplus_{l=0}^{2} D^{l}(r)\right] Q, \tag{1}$$

- The symmetric and anti-symmetric parts of A remain symmetric and anti-symmetric respectively under rotations.
- The 6-dimensional space can be further broken down, because scalar matrices $A_{ij} = \alpha\delta_{ij}$ and traceless symmetric matrices also transform independently. Thus a rank-2 tensor decomposes into representations of dimension 1 (trace), 3 (anti-symmetric part), and 5 (traceless symmetric part).
- In representation-theoretic terms, we have reduced the 9-dimensional representation ρ into irreducible representations of dimension $1, 3$ and 5.

Steerable filter κ

A filter κ is rotation-steerable if

- it is a normal convolution (cross correlation).
- And satisfying the constraint

$$\kappa(rx) = \rho_2(r)\kappa(x)\rho_1(r)^{-1}. \tag{2}$$

Steerable filter space

Steerable filters form a subspace of the 3D convolution space

- the K_n-dimensional feature vectors $f(x) = \oplus_i f^i(x)$ consist of irreducible features $f^i(x)$ of dimension $2\, l_{in} + 1$.
- $\kappa : \mathbb{R}^3 \to \mathbb{R}^{K_{n+1} \times K_n}$ splits into blocks $\kappa^{jl} : \mathbb{R}^3 \to \mathbb{R}^{(2j+1)\times(2l+1)}$ mapping between irreducible features.

$$\kappa^{jl}(rx) = D^j(r)\kappa^{jl}(x)D^l(r)^{-1}. \tag{3}$$

Basic matrix of the steerable filter space

$$\mathrm{vec}(\kappa^{jl}(rx)) = [D^j \otimes D^l](r)\,\mathrm{vec}(\kappa^{jl}(x)), \tag{4}$$

$$[D^j \otimes D^l](r) = Q^T \left[\bigoplus_{J=|j-l|}^{j+l} D^J(r)\right] Q \tag{5}$$

Thus, we can change the basis to $\eta^{jl}(x) := Q\,\mathrm{vec}(\kappa^{jl}(x))$ such that constraint 3 becomes

$$\eta^{jl}(rx) = \left[\bigoplus_{J=|j-l|}^{j+l} D^J(r)\right] \eta^{jl}(x). \tag{6}$$

$$\eta^{jl}(x) = \bigoplus_{J=|j-l|}^{j+l} \eta^{jl,J}(x)\ , \qquad \eta^{jl,J}(rx) = D^J(r)\eta^{jl,J}(x) \tag{7}$$

Basic matrix of the steerable filter space

A famous equation for which the *unique* and *complete* solution is well-known to be given by the spherical harmonics $Y^J(x) = (Y^J_{-J}(x), \ldots, Y^J_J(x)) \in \mathbb{R}^{2J+1}$.

$$\eta^{jl,Jm}(x) = \varphi^m(\|x\|)\, Y^J(x/\|x\|) \tag{8}$$

$$\varphi^m(\|x\|) = \exp\left(-\frac{1}{2}(\|x\| - m)^2/\sigma^2\right) \tag{9}$$

Tetris recognition proposed in [9]

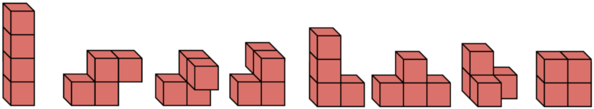

Figure 2: 3D Tetris shapes. Blocks correspond to single points. The third and fourth shapes from the left are mirrored versions of each other.

- Task: classifying 8 kinds of Tetris blocks (voxel grids), in a fixed orientation
- Model: 4-layer 3D Steerable CNN vs conventional CNN
- : Performance: $99 \pm 2\%$ vs $27 \pm 7\%$
- It seems [9] did NOT present the result on the task.

3D model classification

	micro			macro			total		
	P@R	R@N	mAP	P@R	R@N	mAP	score	input size	params
Furuya [5]	**0.814**	0.683	0.656	**0.607**	0.539	**0.476**	**1.13**	126×10^3	8.4M
Esteves [4]	0.717	0.737	0.685	0.450	0.550	0.444	**1.13**	$\mathbf{2 \times 64^2}$	0.5M
Tatsuma [8]	0.705	**0.769**	**0.696**	0.424	**0.563**	0.418	1.11	38×224^2	3M
Ours	0.704	0.706	0.661	0.490	0.549	0.449	1.11	1×64^3	**142k**
Cohen [3]	0.701	0.711	0.676	-	-	-	-	6×128^2	1.4M
Zhou [1]	0.660	0.650	0.567	0.443	0.508	0.406	0.97	50×224^2	36M
Kanezaki [6]	0.655	0.652	0.606	0.372	0.393	0.327	0.93	-	61M
Deng [7]	0.418	0.717	0.540	0.122	0.667	0.339	0.85	-	138M

Table: Results of the SHREC17 experiment.

- Task: classifying 55 classes of 64 x 64 x 64 voxel grids
- Model: 8-layer 3D Steerable CNN

Amino acid environments

- Baseline: either [2] or [10]
- Model: the same dimension in each layer as the baseline but with 3D Steerable CNN
- : Performance: 58% vs 56%

CATH: Protein structure classification

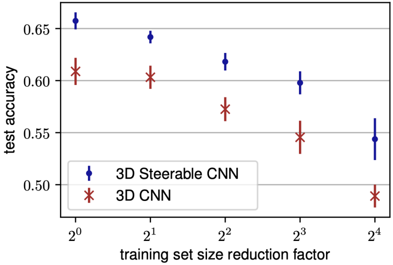

Figure: Accuracy on the CATH test set as a function of increasing reduction in training set size.

- Baseline: ResNet34 with half as many channels as the original (15878764 parameters)
- Model: the same dimension in each layer as the baseline but with 3D Steerable CNN (143560 parameters)

Conclusion

- Not for broad audience
- The math looks solid
- Missing justifications of the engineering choices
- Demonstrate on limited domains

Song Bai, Xiang Bai, Zhichao Zhou, Zhaoxiang Zhang, and Longin Jan Latecki.
Gift: A real-time and scalable 3d shape search engine.
2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 5023–5032, 2016.

Wouter Boomsma and Jes Frellsen.
Spherical convolutions and their application in molecular modelling.
In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 3433–3443. Curran Associates, Inc., 2017.

Taco S. Cohen, Mario Geiger, Jonas Köhler, and Max Welling.
Spherical cnns.
CoRR, abs/1801.10130, 2018.

Carlos Esteves, Christine Allen-Blanchette, Ameesh Makadia, and Kostas Daniilidis.
3d object classification and retrieval with spherical cnns.
Technical report, 2017.

Takahiko Furuya and Ryutarou Ohbuchi.

Deep aggregation of local 3d geometric features for 3d model retrieval.

In Edwin R. Hancock Richard C. Wilson and William A. P. Smith, editors, *Proceedings of the British Machine Vision Conference (BMVC)*, pages 121.1–121.12. BMVA Press, September 2016.

Asako Kanezaki.

Rotationnet: Learning object classification using unsupervised viewpoint estimation.

CoRR, abs/1603.06208, 2016.

M. Savva, F. Yu, Hao Su, M. Aono, B. Chen, D. Cohen-Or, W. Deng, Hang Su, S. Bai, X. Bai, N. Fish, J. Han, E. Kalogerakis, E. G. Learned-Miller, Y. Li, M. Liao, S. Maji, A. Tatsuma, Y. Wang, N. Zhang, and Z. Zhou.

Large-scale 3d shape retrieval from shapenet core55.

In *Proceedings of the Eurographics 2016 Workshop on 3D Object Retrieval*, 3DOR '16, pages 89–98, Goslar Germany, Germany, 2016. Eurographics Association.

Atsushi Tatsuma and Masashi Aono.

Multi-fourier spectra descriptor and augmentation with spectral clustering for 3d shape retrieval.

The Visual Computer, 25:785–804, 2008.

Nathaniel Thomas, Tess Smidt, Steven M. Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley.

Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds.

CoRR, abs/1802.08219, 2018.

Wen Torng and Russ B. Altman.

3d deep convolutional neural networks for amino acid environment similarity analysis.

In *BMC Bioinformatics*, 2017.