

Equivariant Deep Learning via Morphological and Linear Scale Space PDEs on the Space of Positions and Orientations

Remco Duits¹, Bart Smets¹, Erik Bekkers², Jim Portegies¹

¹ Dep. Mathematics and Computer Science, TU/e, the Netherlands,

² Informatics Institute, UvA, the Netherlands

Abstract. We present PDE-based Group Convolutional Neural Networks (PDE-G-CNNs) that generalize Group equivariant Convolutional Neural Networks (G-CNNs). In PDE-G-CNNs a network layer is a set of PDE-solvers where geometrically meaningful PDE-coefficients become trainable weights. The underlying PDEs are morphological and linear scale space PDEs on the homogeneous space \mathbb{M}_d of positions and orientations. They provide an equivariant, geometrical PDE-design and model interpretability of the network.

The network is implemented by morphological convolutions with approximations to kernels solving morphological α -scale space PDEs, and to linear convolutions solving linear α -scale space PDEs. In the morphological setting, the parameter α regulates soft max-pooling over balls, whereas in the linear setting the cases $\alpha = 1/2$ and $\alpha = 1$ correspond to Poisson and Gaussian scale spaces respectively.

We show that our analytic approximation kernels are accurate and practical. We build on techniques introduced by Weickert and Burgeth who revealed a key isomorphism between linear and morphological scale spaces via the Fourier-Cramér transform. It maps linear α -stable Lévy processes to Bellman processes. We generalize this to \mathbb{M}_d and exploit this relation between linear and morphological scale-space kernels.

We present blood vessel segmentation experiments that show the benefits of PDE-G-CNNs compared to state-of-the-art G-CNNs: increase of performance along with a huge reduction in network parameters.

Keywords: Convolutional neural networks, Scale space theory, Cramér transform, Geometric deep learning, Morphological convolutions and PDEs

1 Introduction

Current deep learning with convolutional neural networks (CNNs) [17, 11, 18] performs superbly on classification tasks like vessel segmentations on specific datasets, but requires massive annotated datasets for training because it optimizes inefficiently over huge parameter spaces, and lacks geometric model interpretability because network weights in CNNs have a limited geometric meaning, and it is hard to control or understand the dynamics.

To improve the efficiency of the optimization, one can perform geometric reduction by hard-coding in the system roto-translation equivariance of the network (i.e. roto-translation of input should yield the same roto-translation of output) and roto-translation invariance of classification. To improve the interpretability of the network dynamics, one would like to avoid the use of ad-hoc (yet effective) non-linearities such as ReLU's and max-pooling, as they hamper (stochastic) analysis of the network dynamics via PDEs and central limit theorems.

Recently, inclusion of basic geometric Lie group designs into CNNs has had significant success in group equivariant CNNs (G-CNNs) [7, 20, 24, 4], but lacks a geometric PDE description of crossing-preserving network dynamics on the homogeneous space \mathbb{M}_d of positions and orientations.

We therefore propose a PDE-based equivariant convolutional neural network (PDE-G-CNN) with a linear and morphological scale space design that:

- allows for a serious geometric reduction of network parameters without loss of classification performance (e.g. pixel-wise blood vessel classification).
- allows for a clear geometric analysis of geometric scale space flows through the network, where we build upon our homogeneous space generalization of the core isomorphism [1, 6, 19] of scale space theory: the Cramér transform.

The overall idea is visualized in the context of vessel segmentation in Fig. 1. Here the image data is first lifted to the 3D space $\mathbb{M}_2 \equiv SE(2) := \mathbb{R}^2 \times SO(2)$ of positions and orientations after which we apply PDE-layers strictly involving linear and morphological convolution layers with analytic kernels implementing linear and morphological PDEs on \mathbb{M}_2 . We conclude with a projection layer that integrates over all angles before pixel-wise classification.

.....Our geometric PDEs relate to α -stable Lévy processes [8] and cost-processes akin to [1], but then on \mathbb{M}_d rather than \mathbb{R}^d . This relates to probabilistic equivariant numerical neural networks [14] that use anisotropic convection-diffusions on \mathbb{R}^d . In contrast to these networks, the PDE-G-CNNs that we propose allow for *simultaneous* spatial and angular diffusion on \mathbb{M}_d . Furthermore we include Bellman processes [1] for max pooling over Riemannian balls.

The main contributions of this article are:

- We propose a new PDE-based design of equivariant convolutional neural networks: PDE-G-CNNs where standard nonlinearities are replaced by morphological scale space convolutions.
- We present an (approximate) Cramér transform that generalizes the fundamental isomorphism [1, 6, 19] between linear scale spaces and morphological scale spaces from \mathbb{R}^d towards homogeneous spaces such as \mathbb{M}_d . We use it for new analytic approximation of morphological scale space kernels in Def. 3.
- We assess the quality of the approximative linear scale space kernels in Thm 1 and the quality of the approximative morphological scale space kernels in Prop. 1 with 3 subsequent motivations. We use them in our PDE-G-CNNs.

- We show that PDE-G-CNNs outperform state-of-the-art (awarded) G-CNNs [4] on vessel segmentation while having 10 times less parameters.

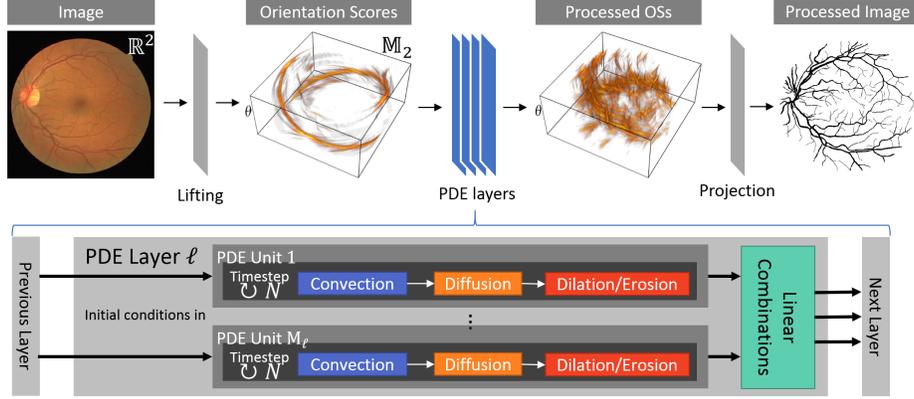


Fig. 1: Illustration of PDE-based CNNs for vessel segmentation. Here convection and diffusion flows (for transport and regularization) will be implemented by linear group convolution, dilation (for max-pooling) and erosion (for data-sharpening) by morphological group convolutions [21].

2 Design of PDE-Based Equivariant Neural Network

2.1 The Lifting Layer: Extending the Image Domain from \mathbb{R}^d to \mathbb{M}_d

In order to disentangle all local orientations in an image we will lift the data from \mathbb{R}^d to the homogeneous space \mathbb{M}_d of positions and orientations given by

$$\mathbb{M}_d = \mathbb{R}^d \rtimes S^{d-1} := G/H, \quad d \in \{2, 3\},$$

with subgroup $H = \{g \in G \mid g \odot \mathbf{p}_0 = \mathbf{p}_0\}$ of roto-translation group $G = SE(d) = \mathbb{R}^d \rtimes SO(d)$, with origin $\mathbf{p}_0 := (\mathbf{0}, \mathbf{a})$ where $\mathbf{a} \in S^{d-1}$ sets an a priori fixed reference axis. Lie group G carries the following group product:

$$g_1 g_2 = (\mathbf{x}_1, \mathbf{R}_1)(\mathbf{x}_2, \mathbf{R}_2) = (\mathbf{x}_1 + \mathbf{R}_1 \mathbf{x}_2, \mathbf{R}_1 \mathbf{R}_2)$$

for all $g_1, g_2 \in G$. The Lie group G acts on $\mathbb{R}^d \rtimes S^{d-1}$ by

$$g \odot \mathbf{p} := (\mathbf{x} + \mathbf{R}\mathbf{x}', \mathbf{R}\mathbf{n}'), \quad \text{for } g = (\mathbf{x}, \mathbf{R}) \in SE(d), \quad \mathbf{p} = (\mathbf{x}', \mathbf{n}') \in \mathbb{R}^d \rtimes S^{d-1}, \quad (1)$$

The case $d = 2$ is of primary interest. Then $H = \{e\}$, with unity element $e = (\mathbf{0}, I)$ so that $\mathbb{M}_2 \equiv SE(2)$. Then we set a priori reference orientation $\mathbf{a} = (1, 0)$ and identify counter-clockwise rotation $\mathbf{R}_\theta \in SO(2)$ with orientation $\mathbf{n}(\theta) = (\cos \theta, \sin \theta) \in S^1$, and angle $\theta \in \mathbb{R}/(2\pi\mathbb{Z})$.

We lift the image data to an orientation score $\mathcal{W}_\psi f : \mathbb{M}_d \rightarrow \mathbb{R}$ that reveals how an image is decomposed out of local orientations. It extends the image domain from \mathbb{R}^d to \mathbb{M}_d and is obtained by convolving the image $f \in \mathbb{L}_2(\mathbb{R}^d)$ by a family of (anisotropic) group-coherent wavelets:

$$\mathcal{W}_\psi f(\mathbf{x}, \mathbf{n}) = \int_{\mathbb{R}^d} \overline{\psi(\mathbf{R}_\mathbf{n}^{-1}(\mathbf{y} - \mathbf{x}))} f(\mathbf{y}) \, d\mathbf{y}, \quad \text{for all } \mathbf{x} \in \mathbb{R}^d, \mathbf{n} \in S^{d-1}, \quad (2)$$

for all roto-translations $g = (\mathbf{x}, \mathbf{R}_\mathbf{n}) \in G$, where $\mathbf{R}_\mathbf{n} \in SO(d)$ is any rotation that rotates reference orientation $\mathbf{a} \in S^{d-1}$ onto \mathbf{n} . See Fig. 1.

In this work we shall train the wavelets $\psi \in \mathbb{L}_2(\mathbb{R}^d)$, but in future work we will show (and motivate) that the trained wavelets allowing for image reconstruction by integration over angles only are remarkably close to cake-wavelets [9].

2.2 PDE Layers by Linear and Morphological Scale Spaces on \mathbb{M}_d

The processing of feature maps on \mathbb{M}_d in our proposed network happens according to the following PDE:

$$\begin{aligned} \frac{\partial W}{\partial t} &= \text{convection} \quad \text{frac. diffusion} \quad \text{dilation/erosion} \\ &= -\mathbf{c} \cdot \nabla_{\mathcal{G}_1} W + |\Delta_{\mathcal{G}_2}|^\alpha W \pm \|\nabla_{\mathcal{G}_3^\pm} W\|^{2\alpha}, \text{ on } \mathbb{M}_d \times \mathbb{R}^+, \\ W|_{t=0} &= \mathcal{W}_\psi f, \quad \text{on } \mathbb{M}_d, \end{aligned} \quad (3)$$

where the orientation score (2) serves as the initial condition, and where the gradient $\nabla_{\mathcal{G}_1} W$ is relative to a left-invariant metric tensor \mathcal{G}_1 on \mathbb{M}_d , and where $\Delta_{\mathcal{G}_2}$ is a left-invariant Laplacian for diffusion indexed by a left-invariant metric tensor \mathcal{G}_2 . For the dilation (+ case) and erosion (- case) part we use metric tensors \mathcal{G}_3^\pm , as we explain below. The convection takes care of equivariant transport, the fractional diffusion for equivariant regularization, the dilation for data-propagation, and the erosion for data-sharpening.

In principle any evolution type PDE can be chosen as the basis of a PDE layer, as long as the PDE is equivariant the resulting layer will also be equivariant. What PDE one chooses would depend on the application. The PDE we propose here is the result of wanting to have behaviour that is known to work well in traditional CNNs, specifically the type of behaviour exhibited by convolution, pooling and ReLU operations. More details on the correspondence between (3) and traditional CNNs can be found in [21].

Geometric parameter reduction: The left-invariant metric must be well-defined on \mathbb{M}_d . This allows for geometric parameter reduction [21, cor.2.7]. E.g. If data-adaptation is omitted one must use [21, Prop.2.8] metric tensors:

$$\mathcal{G}_m|_{\mathbf{p}}(\dot{\mathbf{p}}, \dot{\mathbf{p}}) = \xi^2 |\dot{\mathbf{x}} \cdot \mathbf{n}|^2 + \nu^2 \|\dot{\mathbf{n}}\|^2 + \xi^2 \epsilon_m^{-2} \|\dot{\mathbf{x}} \wedge \mathbf{n}\|^2, \quad (4)$$

for all $\mathbf{p} = (\mathbf{x}, \mathbf{n}) \in \mathbb{M}_d$, and tangent vectors $\dot{\mathbf{p}} = (\dot{\mathbf{x}}, \dot{\mathbf{n}}) \in T_{\mathbf{p}}(\mathbb{M}_d)$, for $m \in \{1, 2, 3\}$. Parameter $\nu > 0$ regulates angular motion costs. By default we set $\nu = 1$ as the cases $\nu > 0$ follow by direct scaling arguments. For $\nu \rightarrow \infty$ and $d = 2$ we arrive at related work [14]. Regarding the anisotropies in \mathcal{G}_m :

- For $m = 1$ (**convection**) we set $\epsilon_1 = \xi = 1$ and only train multiple off-sets \mathbf{c} .
- For $m = 2$ (**diffusion**) and $m = 3$ (**dilation**) we train respectively \mathcal{G}_2 and \mathcal{G}_3^+ . One expects $0 < \epsilon_2, \epsilon_3 \ll 1$ in order to diffuse/dilate primarily along lines.
- For $m = 3$ (**erosion**), i.e. the - case in (3) we train \mathcal{G}_3^- . Here one expects $\xi \gg 1$ and $\xi/\epsilon_3 \ll 1$ so that one sharpens across lines (and not along them).

We will train the metric tensors so that the data determines optimal anisotropies.

Distances on \mathbb{M}_d : The symmetric left-invariant metric tensor fields \mathcal{G}_m on \mathbb{M}_d induces a symmetric left-invariant distance (‘metric’) by

$$d_{\mathcal{G}_m}(\mathbf{p}_0, \mathbf{p}_1) = \inf_{\substack{\gamma(0) = \mathbf{p}_0, \gamma(1) = \mathbf{p}_1, \\ \gamma \in \text{Lip}([0, 1], \mathbb{M}_d)}} \int_0^1 \sqrt{\mathcal{G}_m|_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t))} dt. \quad (5)$$

For all $g \in G = SE(d)$, $\mathbf{p}_1, \mathbf{p}_2 \in \mathbb{M}_d$ one has $d_{\mathcal{G}_m}(\mathbf{p}_1, \mathbf{p}_2) = d_{\mathcal{G}_m}(g \odot \mathbf{p}_1, g \odot \mathbf{p}_2)$.

Logarithmic approximations of distances on \mathbb{M}_d : The metric tensor fields also induce a norm on the tangent space $T_e(G)$ of the unity element $e = (\mathbf{0}, I)$:

$$\|\mathbf{v}\|_{\mathcal{G}} := \left\| \frac{\partial}{\partial t} \exp_G(tv) \odot p_0|_{t=0} \right\|_{\mathcal{G}|_{p_0}} \quad (6)$$

for all $\mathbf{v} \in T_e(G)$, where $\exp_G : T_e(G) \rightarrow G$ is the exp map of G . We define

$$\rho_{\mathcal{G}}(\mathbf{p}) := \inf_{g \in G_{\mathbf{p}}} \|\log_G g\|_{\mathcal{G}} \quad \text{with } G_{\mathbf{p}} := \{g \in G \mid g \odot \mathbf{p}_0 = \mathbf{p}\}. \quad (7)$$

For simple formulas that we used in our code for $d = 2$ see [21]. For formulas for $d = 3$ (identifying the minimum element of $G_{\mathbf{p}}$ above) see [8].

2.3 PDE-Based Deep Learning by G-CNNs on \mathbb{M}_2

Usually CNNs on \mathbb{R}^d [17] iterate 1) possibly off-centered convolution kernels, 2) max-pooling, 3) regularization, 4) ReLU’s. We see them as sampled operator splittings of a PDE evolution combining convection, *fractional* morphology and diffusion on \mathbb{R}^d , with multiple convection-vectors for training center off-sets.

Continuing in this viewpoint on CNNs we replace \mathbb{R}^d by \mathbb{M}_d and obtain equivariant PDE-based CNNs on \mathbb{M}_d . This allows for the geometric reduction of parameters, and we obtain equivariance (w.r.t. translations *and* rotations). Both regular CNNs on \mathbb{R}^2 and equivariant PDE-based CNNs on \mathbb{M}_d are special instances of PDE-based CNNs on homogeneous spaces as shown in Fig. 2.

Recall Fig. 1 where our PDE-based CNNs on \mathbb{M}_d are depicted. At the first layer of the feed forward network we set the orientation score $U = \mathcal{W}_{\psi}f$ as initial condition. In subsequent layers we use the output of the previous layer as the initial conditions of the PDEs and take the solutions of the PDEs at a fixed time $t = T$ to be used for further processing by batch-normalization and linear combinations. The resulting maps are then passed to the next layer and we repeat until we have gone through all the layers. The choice $t = T$ is not crucial as the α -homogeneity of the generators in PDE (3) ensures time-scale invariance.

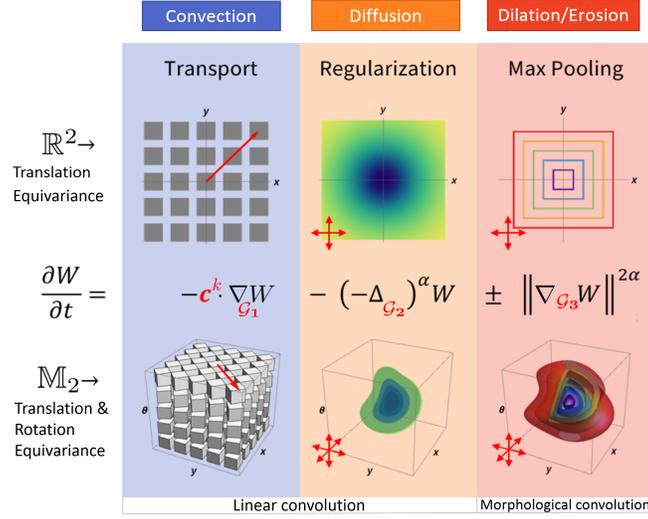


Fig. 2: Geometric design of neurons by PDEs in homogeneous space \mathbb{R}^2 (top) and in homogeneous space \mathbb{M}_2 (bottom), the trainable parameters are highlighted in red. Transport and regularization are solved by linear convolutions (8), and dilation/erosion are solved by morphological convolutions (9). For dilation with $\alpha = \frac{1}{2}$ this boils down to max pooling over geometric balls that grow with $t > 0$, as depicted bottom right.

3 Linear and Morphological Kernel Implementation

The operator splitting of the equivariant PDE evolution (3) as depicted in Fig. 2, boils down to iteratively activating one of the terms in their generator.

Firstly, the convection PDE part in Fig. 2 is solved by a transport along an exponential curve, for details see [21, Prop.5.1].

Secondly, fractional diffusion in (3) is solved by a linear group convolution:

$$W(\mathbf{p}, t) = (K_t^\alpha * U)(\mathbf{p}) = \int_G K_t^\alpha(g^{-1} \odot \mathbf{p}) U(g \odot \mathbf{p}_0) dg \quad (8)$$

with α -scale space kernel K_t^α , where $\mathbf{p}_0 = (\mathbf{0}, \mathbf{a})$ is the origin in the homogeneous space \mathbb{M}_d using the action (1) of Lie group $G = SE(d)$, and Haar measure dg .

Thirdly, viscosity solutions [13] of dilations/erosions are morphological convolutions:

$$W(\mathbf{p}, t) = (k_t^\alpha \square U)(\mathbf{p}) = \pm \inf_{g \in G} \{k_t^\alpha(g^{-1} \odot \mathbf{p}) \pm U(g \odot \mathbf{p}_0)\}. \quad (9)$$

Here the kernels k_t^α are positive and the $-$ cases above solve the dilation PDE, and the $+$ cases solves the erosion PDE, cf. (3). Tangible approximations for the kernels will be obtained further on by approximations of the Cramér transform.

The next subsections (3.1 and 3.2) are necessarily technical. The main message is that linear kernels K_t^α in (8) are well-approximated by $K_t^{\alpha, appr}$ (Defs. 1 and 2), whereas morphological kernels k_t^α in (9) are well-approximated by $k_t^{\alpha, appr}$ (Def. 3).

3.1 Analytic Approximations of α -scale space kernels on \mathbb{M}_2

For exact solutions of fractional diffusion kernels see [8]. For more tangible analytic approximations for $\alpha = 1$ see [9]. They are due to a general theory of estimating heat kernels [5, Thm 3.1], [16, thm 7.6, p.53], [12] on Riemannian manifolds and allow us to provide more explicit bounds. First we summarize these existing mathematical results for $\alpha = 1$ on $(\mathbb{M}_2, \mathcal{G}_2)$ and then generalize, to the fractional diffusion case where $\alpha \in (0, 1]$.

Lemma 1. *Consider Riemannian manifold $(\mathbb{M}_2, \mathcal{G} = \mathcal{G}_2)$ with metric $d_{\mathcal{G}}$. Let $M_{opt} := \mathbb{M}_2 \setminus \text{cut}(p_0)$ denote the complement of the closure of the cut-locus. Then there exists $c_j \in C^\infty(M_{opt})$ with bounded partial derivatives s.t.*

$$K_t^1(\mathbf{p}) = \frac{e^{-\frac{d_{\mathcal{G}}(\mathbf{p}, \mathbf{p}_0)^2}{4t}}}{(4\pi t)^{\frac{3}{2}}} \left(\sum_{j=0}^N c_j(\mathbf{p}) t^j + O(t^{N+1}) \right) \quad (10)$$

for all $\mathbf{p} \in M_{opt}$, where $N \geq 0$, and where we may approximate:

$$d_{\mathcal{G}}(\mathbf{p}, \mathbf{p}_0) = \rho_{\mathcal{G}}(\mathbf{p}) + \epsilon(\mathbf{p}), \quad (11)$$

with $\epsilon(\mathbf{p}) \rightarrow 0$ quadratically and $c_0(\mathbf{p}) \rightarrow 1$ for $\mathbf{p} \rightarrow \mathbf{p}_0$. For $N = 0$ we have:

$$\begin{aligned} K_t^1(\mathbf{p}) &= \frac{e^{-\frac{(d_{\mathcal{G}}(\mathbf{p}, \mathbf{p}_0))^2}{4t}}}{(4\pi t)^{\frac{3}{2}}} \left(c_0(\mathbf{p}) + O\left(\frac{t}{(\rho(\mathbf{p}))^2}\right) \right) \text{ for } 0 < \frac{\sqrt{t}}{\rho(\mathbf{p})} \ll 1, \\ \exists C \geq 1, D \geq 1 \forall t > 0 : C^{-1} \frac{e^{-\frac{(\rho(\mathbf{p}))^2}{4D^{-1}t}}}{(4\pi D^{-1}t)^{\frac{3}{2}}} &\leq K_t^1(\mathbf{p}) \leq C \frac{e^{-\frac{(\rho(\mathbf{p}))^2}{4Dt}}}{(4\pi Dt)^{\frac{3}{2}}}, \end{aligned} \quad (12)$$

for $\mathbf{p} \in \mathcal{C}$, with $\mathcal{C} \subset M_{opt}$ compact around \mathbf{p}_0 , where we set $\rho(\mathbf{p}) = \rho_{\mathcal{G}}(\mathbf{p})$.

Remark 1. (anisotropy) Above c_0 is slowly varying iff Ricci curvatures are bounded, holding for finite anisotropy ϵ_2^{-1} in \mathcal{G}_2 , recall (4). In normalized coordinates $\{y^i\}$ one has: $c_0(\mathbf{p}) = 1 - \frac{1}{12} \sum_{i=1}^3 R_{ii}(y^i(\mathbf{p}))^2 + O(\|y(\mathbf{p})\|^3)$ with $R_{11} = -R_{33} = \frac{1}{2}(\epsilon_2^2 - \epsilon_2^{-2})$, $R_{22} = -\frac{1}{2}(\epsilon_2 - \epsilon_2^{-1})^2$, see [23, Ex.2.7] for details.

Remark 2. (assumption) We assume that $\mathcal{G} = \mathcal{G}_2$ is chosen such that the exact kernels are concentrated on $\mathcal{C} \subset M_{opt} \subset \tilde{\mathbb{M}}_2 := \mathbb{R}^2 \times [-\pi, \pi]$. Then by (4) we have constant $\det(\mathcal{G}) = \xi^4 \nu^2 \epsilon_2^{-2}$ and

$$1 = \int_{\mathbb{M}_2} K_t^1(x, y, \theta) \sqrt{\det \mathcal{G}} \, dx dy d\theta \approx \int_{\mathcal{C} \subset M_{opt}} K_t^1(x, y, \theta) \sqrt{\det \mathcal{G}} \, dx dy d\theta. \quad (13)$$

This assumption makes estimates (12) meaningful and avoids periodicity hassles.

Lemma 1 motivates the following approximation kernel for $\alpha = 1$.

Definition 1. *Now we define the approximate Gaussian kernel*

$$K_t^{1,appr}(\mathbf{p}) := \frac{J(\mathbf{p})}{(4\pi t)^{3/2}} e^{-\frac{(\rho_{\mathcal{G}}(\mathbf{p}))^2}{4t}}, \mathbf{p} \in \mathbb{M}_2, t > 0. \quad (14)$$

with Jacobian $J(\mathbf{p}) := \frac{(\theta/2)^2}{\sin^2(\theta/2)}$ for $\mathbf{p} = (\mathbf{x}, \theta) \in \mathbb{M}_2$, and $J(\mathbf{x}, 0) = 1$.

Regarding the Jacobian we note that the logarithm on $G = SE(2)$ equals [21]:

$$\text{Log}_G(\mathbf{p}) = \left(\frac{\theta}{2} \left(y + x \cot \left(\frac{\theta}{2} \right) \right), \frac{\theta}{2} \left(-x + y \cot \left(\frac{\theta}{2} \right) \right), \theta \right), \quad \theta \neq 0, \quad (15)$$

for all $\mathbf{p} = (x, y, \theta)$. Thereby $J(\mathbf{p}) = |\det D_{\mathbf{p}} \text{Log}_G(\mathbf{p})|$, and by assumption in Remark 2 kernel $K_t^{1,appr}$ is nearly \mathbb{L}_1 -normalised w.r.t. Haar-measure $\sqrt{\det \mathcal{G}} dx dy d\theta$.

From semi-group theory [25] it follows that semi-groups generated by taking fractional powers of the generator (in our case $\Delta_G \mapsto -|\Delta_G|^\alpha$) amount to the following **key relation** between the α -kernel and the diffusion kernel:

$$K_t^\alpha(\mathbf{p}) := \int_0^\infty q_{t,\alpha}(\tau) K_\tau^1(\mathbf{p}) d\tau, \quad \text{for } \alpha \in (0, 1], t > 0. \quad (16)$$

where $q_{t,\alpha} \geq 0$ is the inverse Laplace transform of $r \mapsto e^{-tr^\alpha}$, see [25].

Definition 2. *Akin to (16) we set for $0 < \alpha < 1, t > 0$ the kernel*

$$K_t^{\alpha,appr}(\mathbf{p}) := \int_0^\infty q_{t,\alpha}(\tau) K_\tau^{1,appr}(\mathbf{p}) d\tau \geq 0, \quad \mathbf{p} \in \mathbb{M}_2. \quad (17)$$

Their \mathbb{L}_1 -norms satisfy $\|K_t^\alpha\| = \int_0^\infty \|K_\tau^1\| d\tau \approx \int_0^\infty q_{t,\alpha}(\tau) 1 d\tau = e^{-tr^\alpha}|_{r=0} = 1$.

Next we generalize Lemma 1 and show that $K_t^{\alpha,appr} \approx K_t^\alpha$, for all $\alpha \in (0, 1]$:

Theorem 1. *The approximative Poisson kernel on \mathbb{M}_2 given by (17) equals:*

$$\boxed{K_t^{\frac{1}{2},appr}(\mathbf{p}) = \frac{1}{\pi^2} \frac{t J(\mathbf{p})}{(t^2 + \rho(\mathbf{p})^2)^2}} \quad (18)$$

for all $\mathbf{p} \in \mathbb{M}_2, t > 0$, where $\rho(\mathbf{p}) := \rho_{\mathcal{G}}(\mathbf{p})$. For \mathbf{p} in a compact set $\mathcal{C} \subset M_{opt}$:

$$\boxed{\begin{aligned} \exists_{C \geq 1, D \geq 1} \forall_{t > 0} : C^{-1} K_{tD^{-1}}^{\alpha,appr}(\mathbf{p}) &\leq K_t^\alpha(\mathbf{p}) \leq C K_{tD}^{\alpha,appr}(\mathbf{p}), \\ K_t^{\alpha,appr}(\mathbf{p}_0) &= \frac{\Gamma(\frac{3}{2\alpha})}{\alpha \Gamma(\frac{3}{2}) (4\pi t^{\frac{1}{\alpha}})^{\frac{3}{2}}}, \text{ for all } \alpha \in (0, 1]. \end{aligned}} \quad (19)$$

For all $\alpha \in (0, 1]$ the approximation error equals can be estimated by

$$\mathcal{E}^\alpha(\mathbf{p}, t) := |K_t^\alpha(\mathbf{p}) - K_t^{\alpha,appr}(\mathbf{p})| \leq C \frac{1}{4\pi(\rho(\mathbf{p}))^2} K_t^{\alpha,1D}(\rho(\mathbf{p})) \quad (20)$$

for $C > 0$ and t/ρ^2 small. Here $K_t^{\alpha,1D}$ is the α -stable Lévy process kernel on \mathbb{R} .

For $0 < \alpha \leq \frac{1}{2}$ we have asymptotic expansion for $\frac{t^{2\alpha}}{\rho(\mathbf{p})} \ll 1$:

$$\mathcal{E}^\alpha(\mathbf{p}, t) \equiv \frac{-2\alpha}{4\pi^2(\rho(\mathbf{p}))^3} \sum_{m=0}^{\infty} \sin((m+1)\pi\alpha) \left(\frac{-t}{\rho(\mathbf{p})^{2\alpha}} \right)^{m+1} \frac{\Gamma((m+1)2\alpha)}{m!} \sim O(t). \quad (21)$$

For $\frac{1}{2} \leq \alpha \leq 1$ we also have $\mathcal{E}^\alpha(\mathbf{p}, t) \sim O(t)$,

Proof. First we derive (18). From Yosida's computations on (16), using contour integration in \mathbb{C} for the inverse Laplace transform for $q_{t, \frac{1}{2}}$ one has on \mathbb{R}^n :

$$\int_0^\infty q_{t, \alpha = \frac{1}{2}}(\tau) \frac{e^{-\frac{\rho^2}{4\tau}}}{(4\pi t)^{\frac{n}{2}}} d\tau = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\pi^{\frac{n+1}{2}}} \frac{t}{(t^2 + r^2)^{\frac{n+1}{2}}}$$

with radius $r = \|\mathbf{x}\| > 0$, $\mathbf{x} \in \mathbb{R}^n$. Now by Def. 2 we must essentially redo the same computations but by Def. 1 we must now substitute $r = \rho_{\mathcal{G}}(\mathbf{p})$ $n = \dim(\mathbb{M}_2) = 3$ and multiply by the Jacobian $J(\mathbf{p})$ above. This yields (18).

The top-formula in (19) follows by Lemma 1, $q_{t, \alpha} \geq 0$ and linearity in (16). The formula for $K_t^{\alpha, Appr}(\mathbf{p}_0)$ in (19) comes by inverse Fourier transform:

$$K_t^{\alpha, appr}(\mathbf{p}_0) = \left(\mathcal{F}_{\mathbb{R}^3}^{-1}(\omega \mapsto \frac{1}{(2\pi)^{3/2}} e^{-t\|\omega\|^{2\alpha}}) \right)(\mathbf{0}) = \frac{4\pi}{(2\pi)^3} \int_0^\infty e^{-\left(Rt^{\frac{1}{2\alpha}}\right)^{2\alpha}} R^2 dR$$

with $\|\omega\| = R$. The tangible relation (20) comes by (16) and (17) and Lemma 1:

$$\begin{aligned} |K_t^\alpha(\mathbf{p}) - K_t^{\alpha, Appr}(\mathbf{p})| &= \int_0^\infty q_{t, \alpha}(\tau) |K_\tau^1(\mathbf{p}) - K_\tau^{1, Appr}(\mathbf{p})| d\tau \\ &\stackrel{\text{Lemma 1}}{\leq} C \int_0^\infty q_{t, \alpha}(\tau) \frac{\tau}{(\rho(\mathbf{p}))^2} K_\tau^{1, appr}(\mathbf{p}) d\tau \text{ for } t/\rho^2 \text{ small} \\ &= \frac{C}{4\pi(\rho(\mathbf{p}))^2} \int_0^\infty q_{t, \alpha}(\tau) \frac{(4\pi\tau)}{(4\pi\tau)^{\frac{3}{2}}} e^{-\frac{\rho^2(\mathbf{p})}{4\tau}} d\tau = C \frac{K_t^{\alpha, 1D}(\rho(\mathbf{p}))}{4\pi(\rho(\mathbf{p}))^2}. \end{aligned}$$

For (21) we rely on (20) and asymptotic formulas for the 1D α -stable Lévy process kernels in [15, Eq.2.3]. The order term for the cases $1 \geq \alpha > \frac{1}{2}$ follows by similar asymptotic behavior [15, Eq.2.3] for 1D α -kernels when $x = \rho/t^{\frac{1}{2\alpha}} \rightarrow \infty$ \square

From [9, ch:5.4.1] it follows that one can set $(C, D) = (1, 2)$. If \mathcal{G}_2 is spatially isotropic ($\epsilon_2 = 1$, one has $(C, D) \approx (1, 1)$, for t small, recall Remarks 1 and 2.

3.2 Analytic Approximations of α -Dilation/Erosion kernels on \mathbb{M}_2

For functions on \mathbb{M}_2 that have the property that the Fourier transform $\mathcal{F}[\mathbf{v} \mapsto f((\exp_G \mathbf{v}) \odot \mathbf{p}_0)]$ is real-valued and non-negative for all $\mathbf{v} \in T_e G \equiv \mathbb{R}^3$, $G = SE(2)$, we define the approximate Cramér-Fourier transform as

$$\mathcal{C}^{appr} f(\mathbf{p}) := \mathcal{C}[\mathbf{v} \mapsto f((\exp_G \mathbf{v}) \odot \mathbf{p}_0)](\log_G g_{\mathbf{p}}), \quad (22)$$

where \exp_G and \log_G , recall (15), denote the exponential and logarithm in $G = SE(2)$, where $\mathbf{p} = (\mathbf{x}, \theta) = g_{\mathbf{p}} \odot \mathbf{p}_0 \in \mathbb{M}_2$ with $g_{\mathbf{p}} = (\mathbf{x}, \mathbf{R}_\theta) \in SE(2)$. Here \mathcal{C} is the Fourier-Cramér transform on $T_e(G) \equiv \mathbb{R}^3$ given in [19]. It concatenates Fourier transform \mathcal{F} , logarithm and Fenchel transform so that $\mathcal{C}(f * g) = \mathcal{C}f \square \mathcal{C}g$.

Definition 3. Let $\rho_{\mathcal{G}_3}$ be the estimate (7) of the Riemannian distance (5) between \mathbf{p} and \mathbf{p}_0 induced by metric tensor field \mathcal{G}_3 (4). Then we define

$$k_t^{\alpha, appr}(\mathbf{p}) := c_\alpha \left(\frac{(\rho_{\mathcal{G}_3}(\mathbf{p}))^{2\alpha}}{t} \right)^{\frac{1}{2\alpha-1}}, \quad k_t^{1/2, appr}(\mathbf{p}) := \begin{cases} 0 & \text{if } \rho_{\mathcal{G}_3}(\mathbf{p}) \leq t, \\ \infty & \text{if } \rho_{\mathcal{G}_3}(\mathbf{p}) > t, \end{cases} \quad (23)$$

with $c_\alpha = \frac{2\alpha-1}{(2\alpha)^{2\alpha/(2\alpha-1)}}$ for $\frac{1}{2} < \alpha \leq 1$.

Proposition 1. *The approximate morphological convolution kernels relate to the linear convolution kernels as follows: $\mathcal{C}^{appr} K_t^{\alpha,appr} = k_t^{\alpha,appr}$.*

Proof. The right-hand side of (23) is the pullback of the morphological kernel on \mathbb{R}^3 via the isomorphism of Lie-algebra $T_e(G)$ and \mathbb{R}^3 given by $c^1 \partial_x|_e + c^2 \partial_y|_e + c^3 \partial_\theta|_e \leftrightarrow (c^1, c^2, c^3)$. Then by [19, Tab.2] one has $\mathcal{C}^{appr} K_t^{\alpha,appr} = k_t^{\alpha,appr}$ \square

We have 3 main motivations for the kernel approximations in Definition 3:

1) For $\alpha = \frac{1}{2}$ it follows by Riemannian wavefront propagation on \mathbb{M}_2 [10] that

$$k_t^{1/2}(\mathbf{p}) = \begin{cases} 0 & \text{if } d_{\mathcal{G}_3}(\mathbf{p}, \mathbf{p}_0) \leq t, \\ \infty & \text{if } d_{\mathcal{G}_3}(\mathbf{p}, \mathbf{p}_0) > t. \end{cases} \quad (24)$$

The logarithmic norm *locally* approximates the Riemannian distance, recall (11). Now compare (24) with (23), then for $t > 0$ small there exists $D > 1$ such that

$$k_{tD^{-1}}^{\alpha=1/2,appr} \leq k_t^{\alpha=1/2} \leq k_{tD}^{\alpha=1/2,appr}. \quad (25)$$

For $\alpha = \frac{1}{2}$ the morphological convolutions (9) are ‘max- pooling’ over balls in \mathbb{M}_d . This becomes a ‘soft-max pooling’ for $\alpha \in (\frac{1}{2}, 1]$, see [21, Figs.7&8], where the approximation $k_t^{\alpha,appr}$ is still close enough in practice [21, Fig.10].

2) $k_t^{\alpha,appr}$ is nearly equal to k_t^α if $\epsilon_3 = 1$, $\nu \rightarrow \infty$ or $\xi \rightarrow \infty$ (where the non-commutative group structure is irrelevant allowing us to rely on exact morphological kernels [19] on \mathbb{R}^3), if t is small such that fronts hardly reach $|\theta| = \pi$.

3) The function $\tilde{S}(\mathbf{c}, t) := k_t^{\alpha,appr}(\exp_G(\mathbf{c}))$ solves $\partial_t \tilde{S}(\mathbf{c}, t) = \|\nabla_{\mathbf{c}} \tilde{S}(\mathbf{c}, t)\|^{2\alpha}$, whereas $S(\cdot, t) = k_t^\alpha$ solves $\partial_t S(\mathbf{p}, t) = \|\nabla_{\mathcal{G}_3} S(\mathbf{p}, t)\|^{2\alpha}$ for $t > 0$ and $\mathbf{p} \in \mathbb{M}_2$.

Extensions to $t \geq 0$ require limits of viscosity solutions [2] (‘Lax-Oleinik’ solutions [13]). By the formula of the logarithm (15) and the chain-law one can express $\nabla_{\mathcal{G}_3}$ into the gradient in logarithmic coordinates and after technical asymptotics one has $\|\nabla_{\mathcal{G}_3} S\|^{2\alpha} = \|(\log_{\mathcal{G}_3} \tilde{S})^{-1}\|^{2\alpha} (1 + O(\alpha\nu^{-2}(1 - (\epsilon_3)^2)t^{1/\alpha}\xi^{-2}))$. This reveals the nearly exact cases in 2) and supports our *conjecture* that (25) generalizes to $\alpha \in [\frac{1}{2}, 1]$, which may be shown via comparison principles [2].

Ad-hoc nonlinearities in CNNs are obsolete and excluded in our PDE-G-CNNs:

Theorem 2. *Let $f : \mathbb{M}_d \rightarrow \mathbb{R}$ be continuous and compactly supported. Let $\emptyset \neq S \subset \mathbb{M}_d$. Let $M = \max_{\mathbf{y} \in \mathbb{M}_d} f(\mathbf{y})$. Define $k_S, k^M : \mathbb{M}_d \rightarrow \mathbb{R} \cup \{\infty\}$ by:*

$$k_S(\mathbf{p}) := \begin{cases} 0 & \text{if } \mathbf{p} \in S, \\ \infty & \text{else,} \end{cases} \quad k^M(\mathbf{p}) := \begin{cases} 0 & \text{if } \mathbf{p} = \mathbf{p}_0, \\ M & \text{else.} \end{cases} \quad (26)$$

Then max-pooling and ReLU’s boil down to morphological convolutions:

$$\sup_{g \in G: g^{-1} \odot \mathbf{p} \in S} f(g \odot \mathbf{p}_0) = -(k_S \square - f)(\mathbf{p}), \quad \max\{0, f(\mathbf{p})\} = -(k^M \square - f)(\mathbf{p})$$

Proof. Follows by computation of (9); for details see [21, Prop.5.16, 5.18]. \square

4 Experimental Observations and Analysis

We test the viability of PDE-G-CNNs for automatic segmentation of vasculature, recall Fig. 1. To this end we consider a roto-translation equivariant CNN on \mathbb{M}_2 where convolution layers involve split-operator solvers (as depicted in Fig. 1) solving the following evolution equation:

$$\frac{\partial W}{\partial t} = -\mathbf{c} \cdot \nabla W + \|\nabla_{\mathcal{G}_3^+} W\|^{2\alpha} - \|\nabla_{\mathcal{G}_3^-} W\|^{2\alpha},$$

wherein we set $\alpha = 0.65$ and train the convection vectors \mathbf{c} and Riemannian metric tensor fields \mathcal{G}_3^- and \mathcal{G}_3^+ (4). Since the PDE contains a convection, a dilation and an erosion term we refer to this layer as a *CDE* layer. The convection is implemented by linear interpolation, the dilation and erosion is implemented with morphological convolution (9) with kernels from Def. 3. A batch normalization and linear combination step complete each layer, see Fig. 1.

The PDE-based operators are implemented as an extension to the PyTorch deep learning framework. Training is performed by standard back-propagation using the ADAM stochastic gradient descent algorithm with a continuous DICE loss function. The same learning rate is used for all parameters and a small amount of L^2 regularization is added to the loss. The choice for $\alpha = 0.65$ has been made experimentally.

We consider three networks: a traditional CNN, a Group CNN and a PDE-G-CNN and apply these to the retinal vessel segmentation problem, specifically the publicly available DRIVE dataset[22], where we classify each pixel of the input image as either being a vessel or background. The accuracy of the model is measured by the average DICE coefficient of the test dataset.

The key properties of these networks, all having 6 layers, are compared in Table 1. Each model was trained 10 times, the distribution of DICE performance per realization is shown in Fig. 3. We observe better performance of our proposed PDE-G-CNNs over G-CNNs [4] even with 10 times less network parameters.

Model	Spatial CNN	Group CNN	(CDE) PDE-G-CNN
Parameters	47352	39258	4128
Test time(s)	1.7	6.5	6.9
Avg. DICE	0.8058	0.8085	0.8115

Table 1: Metrics for the 3 tested models.

Conclusion

We have presented a new PDE-framework for equivariant deep learning. Our PDE-G-CNNs only involve morphological and linear convolutions with scale space kernels. These kernels allow for close analytic approximations (Thm. 1, Def.3) connected by the Cramér transform (22) in Prop.1. Thm.2 and α -stability on \mathbb{M}_d [8] allow us to cut all ad-hoc nonlinearities from CNNs. Practical vessel segmentation experiments show that PDE-G-CNNs outperform state-of-the-art G-CNNs [4] even while reducing the network-complexity by a factor of 10 when using a depth of 6 layers. Ongoing tests by Smets even indicate that the reduction factor increases by a factor of 40 for 12-layer depths, but a full analysis of benefits of PDE-G-CNNs over (G-)CNNs is left for future work.

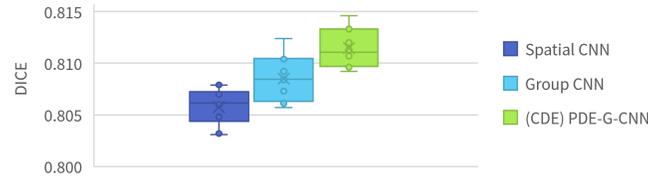


Fig. 3: Distribution of DICE coefficients of the test dataset for the three models.

References

1. Akian, M., Quadrat, J., Viot, M.: Bellman processes. LNCIS **199**, 302–311 (1994)
2. Bardi, M., Capuzzo-Dolcetta, I.: Discontinuous viscosity solutions and applications, pp. 285–358. Birkhauser (1997)
3. Barles, C.: Discontinuous solutions of hamilton-jacobi equations. NATMA **20**(9), 1123–1134 (1993)
4. Bekkers, E., Lafarge, M., Veta, M., Eppenhof, K., Pluim, J., Duits, R.: Roto-translation covariant convolutional networks for medical image analysis. In: MIC-CAI 2018. pp. 440–448 (2018)
5. Ben Arous, G.: Development asymptotique du noyau de la chaleur hypoelliptique sur la diagonale. Ann. Inst. Fourier pp. 73–99 (1989)
6. Burgeth, B., Weickert, J.: An explanation for the logarithmic connection between linear and morphological systems. Proceedings 4th SSVM pp. 325–339 (2003)
7. Cohen, T., Welling, M.: Group equivariant convolutional networks. Proc. of the 33rd Int. Conf. on Machine Learning **48**, 1–12 (2016)
8. Duits, R., Bekkers, E.J., Mashtakov, A.: Fourier transform on \mathbb{M}_3 for exact solutions to linear PDEs. Entropy (SI: 250 year of Fourier) **21**(1), 1–38 (2019)
9. Duits, R., Franken, E.M.: Left invariant parabolic evolution equations on $SE(2)$ and contour enhancement via orientation scores. QAM-AMS **68**, 255–331 (2010)
10. Duits, R., Meesters, S., Mirebeau, J.M., Portegies, J.M.: Optimal Paths of the Reeds-Shepp Car with Applications in image analysis. JMIV **60**(6), 816–848 (2018)
11. Elad, M.: Deep, Deep Trouble. SIAM-NEWS p. 12 (May 2017)
12. ter Elst, A.F.M., Robinson, D.W.: Weighted subcoercive operators on Lie groups. Journal of Functional Analysis **157**, 88–163 (1998)
13. Evans, L.C.: Partial differential equations. AMS, Providence, R.I. (2010)
14. Finzi, M., Bondesan, R., Welling, M.: Probabilistic numeric convolutional neural networks. <https://arxiv.org/pdf/2010.10876.pdf> pp. 1–22 (2020)
15. Garoni, T., Frankel, N.: Lévy flights: Exact results and asymptotics beyond all orders. J. Math. Phys. **43**(5) (2002)
16. Grigorian, A.: Heat Kernel and Analysis on Manifolds. Math. Dep. Bielefeld (2009)
17. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature Research **521**(7553), 436–444 (2015)
18. Litjens, G., Bejnordi, B., van Ginneken, B., Sánchez, C.: A survey on deep learning in medical image analysis. Medical Image Analysis **42**, 60–88 (2017)
19. Schmidt, M., Weickert, J.: Morphological counterparts of linear shift-invariant scale-spaces. Journal of Mathematical Imaging and Vision **56**(2), 352–366 (2016)
20. Siffré, L.: Rigid-Motion Scattering for Image Classification. Ph.D. thesis, Ecole Polytechnique, Paris (2014)
21. Smets, B., Portegies, J., Bekkers, E., Duits, R.: PDE-based group equivariant convolutional neural networks. Tech. rep. (2020), www.arxiv.org/abs/2001.09046

22. Staal, J., Abramoff, M., Niemeijer, M., Viergever, M., van Ginneken, B.: Ridge-based vessel segmentation in images of the retina. *IEEE TMI* **23**(4), 501–509 (2004)
23. Svensson, V.: Curvature of Lie Groups. Math. Dep. of Lund, Sweden (2009)
24. Weiler, M., Geiger, M., Welling, M., Boomsma, W., Cohen, T.: 3D steerable CNNs: Learning equivariant features in volumetric data. *NeurIPS* pp. 1–12 (2018)
25. Yosida, K.: *Functional Analysis*. Springer (1974)