CELLS ARE NOT RECTANGLES: A ROTATION-EQUIVARIANT CNN FOR IMAGES OF THINGS WITH IRREGULAR BOUNDARIES

Anonymous authors
Paper under double-blind review

ABSTRACT

We propose a convolutional neural network (CNN) architecture that is tailored to 2D images of "things" with irregular boundaries, such as: cells, inhomogenous materials, and flatlanders (Abbott, 1884). The CNN is equivariant to $E(2)$, i.e., (continuous) translations, rotations, and reflections in 2D Euclidean space, thus, having features and filters that represent different geometric tensors (i.e., not only scalars, but also quantities related to gradients, elongation, "pointiness", etc.). Each pixel is additionally given a "type", either interior or exterior, and retain this knowledge throughout the convolution layers. Separate convolution filters are learned for passing information within the interior, within the exterior, and in both directions across the interface. To the best of our knowledge, such a fully-equivariant treatment of the boundary of images is new. Moreover, whereas CNNs equivariant to $E(2)$ have already been studied (Weiler & Cesa, 2019; Kondor & Trivedi, 2018), they often not employed in practical situations of relevance, e.g., even in "state-of-the-art" analysis of cell images (e.g., Lafarge et al. (2021); Moen et al. (2019); Ashdown et al. (2020)). We hope that, by describing these ideas in a didactic (and somewhat whimsical) manner, this short format paper can convince more people to use them.

1 A (NOT SO) "CONVOLUTED" ANALOGY (INTRODUCTION)

The remarkable performance of convolutional neural networks for image classification is largely due to the massive reduction in parameters resulting from the use of local convolutional filters. The justification of this "architecturally enforced" assumption stems from two salient properties of natural images:

1. Images are (approximately) *translation-invariant*.
   As such, we want to treat each region of the image as "the same".
   This motivates the use of convolutional filters.
2. Images are (approximately) *local*.
   Thus, the interpretation of a region should depend on its "neighbors".
   This motivates the use of filters that are sparse,
   as they can be zero when far from the pixel in question.

Thus, judicious assumptions about the *structure* of images motivate highly efficient architectures for *analyzing* images.

2 OUR MAIN CONTRIBUTIONS

1. We have developed a neural architecture that is well-adapted for the analysis of biological images. This is achieved by combining general rotational equivariance (an idea that has been described before, but that should be used more often) with a new idea for the treatment of image boundaries.
2. We aim to explain it in a way that is enjoyable to read, and easy to understand, so as to make it fun and accessible to a broader audience.

3. We will perform an "apples-to-apples" comparison (i.e., similar number of parameters and training) of this network with currently-used standard CNNs.
4. We will release the code upon acceptance.

2.1 THE (INCOMPLETE) ELEPHANT IN THE ROOM

Unfortunately, we were unable to provide the code and experimental results for our proposed architecture in time for this submission. Also, the text is clearly rushed. However, we commit to, upon acceptance to the workshop, providing the code for the architecture, and to study and compare its performance with other (*properly matched*) CNNs models[1] in, *at least*, the following real-world dataset of practical importance:

- *Leukemia dataset* (Gupta & Gupta, 2019): Distinguishing between images of normal and cancer cells, which only have subtle morphological differences. The cancers cells are from acute lymphoblastic leukemia – the most common type of childhood cancer (Gupta et al., 2022).

We also commit to making the text in this paper flow much smoother, and include basic examples for understanding the architecture.

3 RELATED WORK

CNNs that are equivariant to $E(2)$, i.e., (continuous) translations, rotations, and reflections in 2D Euclidean space, have been studied in depth by Weiler & Cesa (2019). Moreover, they compared several other approaches that approximately conserved this symmetry, and proposed network architectures that allow for incremental symmetry break along the network.

While there is ever increasing interest in the use of deep neural network architectures to analyse biological images (Kan, 2017; Moen et al., 2019), most of these works do not take fully incorporate the symmetries present in the images in their models, opting instead for pretrained object recognition models (such as "InceptionV3") (Ashdown et al., 2020; Dong et al., 2020; Ramaneswaran et al., 2021).

4 THE CELLNN LAYER (OUR PROPOSED ARCHITECTURE)

There are many cases in which biological data are naturally encoded as 2D images. Often, such images can be considered as having no preferred origin; it could be a small part of a much larger tissue or culture of cells, or the microscope might not have been properly centered around the object of interest. Thus, the well-known architecture of CNNs are an appropriate starting point for analyzing such data. A basic version of such an architecture can be constructed by interleaving convolutional layers with pointwise nonlinearities and pooling/coarsening operations.

4.1 ROTATION EQUIVARIANCE (THE KNOWN IDEA THAT SHOULD BE USED MORE)

In contrast to the images that typically impinge upon our eyes, such microscopic images are less likely to be gravitationally constrained to a preferred orientation. The lack of a preferred orientation calls for an additional symmetry to be incorporated into the architecture — that of 2D rotations. The combination[2] of the 2D translations and 2D rotations is sometimes referred to as $E(2)$, the isometries[3] of 2D Euclidean space.

Without rotational symmetry, standard CNNs have layers which can only assign some number of scalar "features" to each pixel inputed to that layer. To incorporate rotational symmetry,

[1] We deeply believe on the importance of "apples to apples" comparisons; one of the reasons is that such comparisons provide insights even in cases where the theoretical idea does not fulfill the expectations in practice.

[2] semi-direct product

[3] Isometries are maps that preserve the notion of distance.

the analogous features of rotation-equivariant CNNs must be classified by their order.[4] Additionally, the analogous filters of rotation-equivariant CNNs must be classified in the same way. Where standard CNNs simply multiply input features with parameterized filters, rotation-equivariant CNNs must also combine their corresponding representations.[5]

4.2 DOMAIN AWARENESS (THE NEW IDEA THAT SHOULD BE USED FOR CELLS)

When using a standard CNN for rectangular images, pixels near the boundary can often "sense" their position based on how the convolution is performed (e.g., padding with repeated or reflected pixels, periodic boundary conditions, etc.). For objects with irregular boundaries, the typical method is to place it within a bounding box containing constant values for the background. Say, for example, one pads the image with zeros, giving a black background. Then the convolutions will initially consider dark regions in the interior as similar to the background, despite their obvious semantic differences.

In analogy with the fact that physical processes are typically different in different domains (e.g., inside the cytoplasm vs. in the extracellular matrix), the information passed by our convolutions will likewise be domain dependent. To this end, at each layer we premultiply our features with two masks, one highlighting each domain. Different weights are learned for the convolutions applied to each. After convolution, these two masks are applied again to the output, giving a total of four sets of convolutional filters to learn, allowing for separate modelling of signals within and across domain boundaries.

4.3 PARAMETERIZATION FOR THE ROTATION-EQUIVARIANT LAYER

Underlying our "cubist" coarse-graining, one could envision a "platonic" portrait — a continuous function over the 2D plane.As such, we must come to terms with the fact that most rotations do not preserve the orientation of the square grid. Thus, we must settle for an approximate notion of continuous rotational symmetry.

We want our filters to be "steerable". That is, instead of considering nearby pixels in the filter as being "North, South, East, or West" of the center pixel (as in the original (i, j) representation of the images), we want to consider neighboring pixels in the filter as being "Inwards, Outwards, or Sideways". We illustrate how this can be done using the Hermite functions Park et al. (2009) over two variables x_1 and x_2 as an example. These functions are indexed by ordered pairs of nonnegative integers p_1 and p_2, and have rotational orders $P = p_1 + p_2$. There are six Hermite functions of order at most $P = 2$ (Fig. 1):

$$f_{00}(x_1, x_2) \propto e^{-\frac{1}{2}(x_1^2+x_2^2)} \tag{1}$$
$$f_{10} \propto x_1 f_{00} \tag{2}$$
$$f_{01} \propto x_2 f_{00} \tag{3}$$
$$f_{11} \propto x_1 x_2 f_{00} \tag{4}$$
$$f_{20} \propto \left(2x_1^2 - 1\right) f_{00} \tag{5}$$
$$f_{02} \propto \left(2x_2^2 - 1\right) f_{00} \tag{6}$$

Loosely, these functions can be thought of as weighted estimates of derivatives with respect to x_1 and x_2,[6] where the weight is a centered Gaussian with unit variance.

REFERENCES

Edwin Abbott. *Flatland: A Romance of Many Dimensions*. Seeley & Co., 1884.

[4] More specifically, their representation. In the case of 2D rotations, these representations are conveniently indexed by the integers.

[5] For other groups (such as 3D rotations), such operations are highly non-trivial. However, for 2D rotations, representations conveniently combine by adding their integer-valued orders.

[6] Eg, for a function $z(x_1, x_2)$, $\partial z/\partial x_1$ is proportional to the integral of $f_{10}z$ over $\mathbb{R}^2$, aka, their "inner product" $\langle f_{10}|z\rangle$. Likewise, the integral of $f_{11}z$ is proportional to $\partial^2 z/\partial x_1 \partial x_2$.

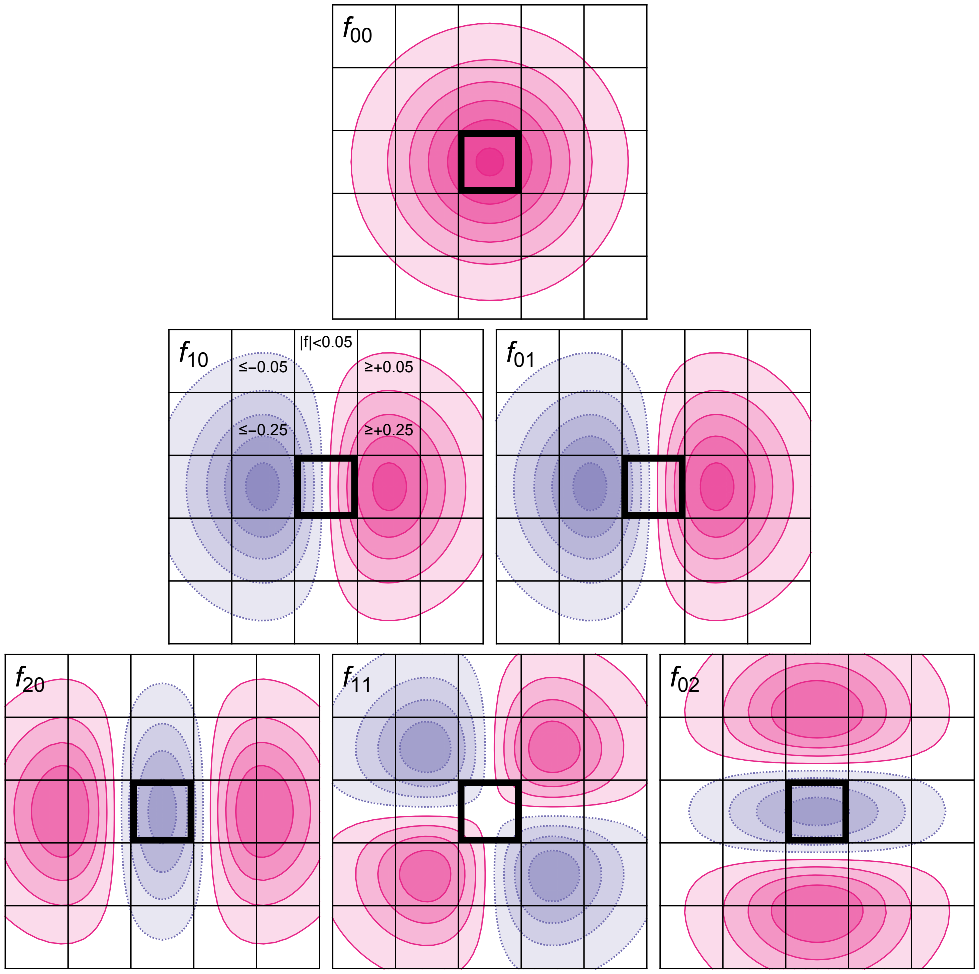

Figure 1: **The 2D Hermite functions of at most second order.**
The first row shows the single zeroth-order Hermite function, a 2D spherical Gaussian. The second row shows the two first-order Hermite functions, and the third row shows the three second-order ones. In all figures, contours denote changes of 0.1, where the white interval is centered around 0 (**pink** indicates **positive**, and **purple negative**). The grid lines denote the pixel boundaries of our 5×5 example filter.

George W Ashdown, Michelle Dimon, Minjie Fan, Fernando Sánchez-Román Terán, Kathrin Witmer, David CA Gaboriau, Zan Armstrong, D Michael Ando, and Jake Baum. A machine learning approach to define antimalarial drug action from heterogeneous cell-based screens. *Science advances*, 6(39):eaba9338, 2020.

Na Dong, Li Zhao, Chun-Ho Wu, and Jian-Fang Chang. Inception v3 based cervical cell classification combined with artificially extracted features. *Applied Soft Computing*, 93:106311, 2020.

Anubha Gupta and Ritu Gupta. ALL Challenge dataset of ISBI 2019 [Data set]. The Cancer Imaging Archive, 2019. URL https://doi.org/10.7937/tcia.2019.dc64i46r.

Ritu Gupta, Shiv Gehlot, and Anubha Gupta. C-NMC: B-lineage acute lymphoblastic leukaemia: A blood cancer dataset. *Medical Engineering & Physics*, 103:103793, 2022.

Andrey Kan. Machine learning applications in cell image analysis. *Immunology and cell biology*, 95(6):525–530, 2017.

Risi Kondor and Shubhendu Trivedi. On the generalization of equivariance and convolution in neural networks to the action of compact groups. In *International Conference on Machine Learning*, pp. 2747–2755. PMLR, 2018.

Maxime W Lafarge, Erik J Bekkers, Josien PW Pluim, Remco Duits, and Mitko Veta. Roto-translation equivariant convolutional networks: Application to histopathology image analysis. *Medical Image Analysis*, 68:101849, 2021.

Erick Moen, Dylan Bannon, Takamasa Kudo, William Graf, Markus Covert, and David Van Valen. Deep learning for cellular image analysis. *Nature methods*, 16(12):1233–1246, 2019.

Wooram Park, Gregory Leibon, Daniel N Rockmore, and Gregory S Chirikjian. Accurate image rotation using Hermite expansions. *IEEE transactions on image processing*, 18(9):1988–2003, 2009.

S Ramaneswaran, Kathiravan Srinivasan, PM Durai Raj Vincent, and Chuan-Yu Chang. Hybrid inception v3 xgboost model for acute lymphoblastic leukemia classification. *Computational and Mathematical Methods in Medicine*, 2021:1–10, 2021.

Maurice Weiler and Gabriele Cesa. General e(2)-equivariant steerable cnns. *Advances in Neural Information Processing Systems*, 32, 2019.

A Appendix (The Telomeres)

A.1 Hermite Functions

Evaluating the Hermite functions at the neighboring vertices appears to produce vectors that remain nearly orthonormal (Fig. 2). If one truly needs orthonormal basis vectors, the Gram-Schmidt orthonormalization procedure would not significantly disturb their interpretability. By the same token, such a step is not likely to be necessary in practice.

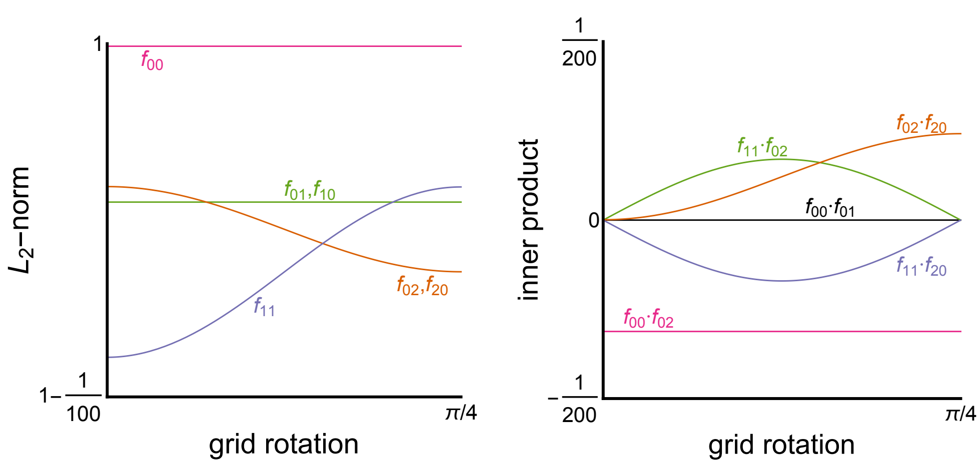

Figure 2: **Discretized Hermite functions are nearly orthonormal.**
We parameterize our rotation-equivariant convolution filters using the Hermite functions. While the (continuous) Hermite functions are orthonormal over $\mathbb{R}^2$, their evaluation at the neighboring $(2S-1)^2$ points may not be. Here, we test how close these vectors are to orthonormal. Even with a (seemingly coarse) 5×5 grid, the *left* figure shows that the resulting vectors remain nearly unit length. Moreover, the *right* figure shows that these vectors also remain nearly orthogonal.