Hyperbolic Equivariant Convolutional Neural Networks for Fish-Eye Image Processing

Pierre-Yves Lagrave, Frédéric Barbaresco

Hyperbolic Equivariant Convolutional Neural Networks for Fish-Eye Image Processing

Pierre-Yves Lagrave[1][0000−0002−5774−636X] and Frédéric Barbaresco[2][0000−0003−3664−3609]

[1] Thales Research and Technology, Palaiseau, France
pierre-yves.lagrave@thalesgroup.com
[2] Thales Land and Air Systems, Limours, France
frederic.barbaresco@thalesgroup.com

Abstract. Fish-Eye image processing with conventional Machine Learning algorithms such as Convolutional Neural Networks is a challenging task because of the distortion effects induced by dewarping the raw hemispherical images into the Euclidean plane. We introduce in this paper an approach based on the emerging field of Geometric Deep Learning in which hyperbolic projection techniques are coupled with equivariance mechanisms in order to preserve native geometrical dependencies and to achieve robustness with respect to natural variations in the perception of the Fish-Eye images. This work in particular motivates the development of efficient $SU(1,1)$ and $SL(2,\mathbb{R})$ Equivariant Neural Networks.

Keywords: Fish-Eye image · Hyperbolic Geometry · Equivariant Neural Networks · Geometrical Robustness · Lie Groups.

1   Introduction and Motivations

Fish-Eye lenses [3] are hemispherical sensors (180° field-of-view) and the native geometry of the corresponding images is therefore spherical, leading to distortion effects once projected onto the 2D Euclidean plane, as illustrated in Figure 1. Using conventional Deep Learning algorithms such as Convolutional Neural Networks (CNN) [22] for Fish-Eye image processing therefore does not appear well suited and other approaches have to be envisioned for this purpose.

More precisely, the problem of Fish-Eye image processing with Deep Learning algorithms is getting more and more traction as Fish-Eye sensors are being used for practical applications, such as perception tasks for autonomous driving [26,30], visual-inertial odometry [15,23] and drone operation [24], with 3 main families of approaches being considered:

– Adjustment of usual algorithms to take into account distortion effects during the learning phase performed on the projected images, including the use of data-augmentation techniques [27,29,13] and that of deformable convolution operators [9,25]

Fig. 1. First high-resolution color image sent by Hazard Cameras (Hazcams) of NASA's Perseverance Mars vehicle after its landing on February 18, 2021. Distortion effects are in particular visible through the curved horizon line. Photo: NASA / JPL-Caltech.

– Transfer of the algorithms learned on usual planar images by modifying the algorithmic characteristics (e.g., weights, convolution kernels, etc.) a posteriori [28,8]
– Using algorithms operating directly in spherical geometry for omnidirectional images (360° field-of-view), as in [14]. In this context, spherical CNN [7,10,17] achieve equivariance with respect to the 3D rotation group SO(3), the input images being represented as signals on the 2D-sphere. Although operating in the native geometry of the inputs, this approach is however not optimal for Fish-Eye images with hemispherical support.

Although some progress has been made in order to adapt or patch existing architectures to account for the distortion effects, no methodology has been proposed to generalize the translation equivariance property of CNN, which is highly beneficial from both accuracy and robustness standpoints when working with planar images, to the case of Fish-Eye images as it was done with the introduction of spherical neural networks for omnidirectional images.

Hence, we propose in this paper to remedy this point by using projection techniques to represent the native images as signals in the 2D hyperbolic space and then by building neural networks which are equivariant [6] to the group acting on the considered hyperbolic model (e.g., Poincaré Disk $\mathbb{D}$, Poincaré Half-Plane $\mathbb{H}_2$, etc.), therefore generalizing spherical CNN. We do not focus here on the details underlying the building of such neural networks (see for instance [20] for details regarding $SU(1,1)$ Neural Networks for signals on $\mathbb{D}$), but we rather provide a generic framework for Fish-Eye image processing through the combinination of hyperbolic geometry and equivariance mechanisms, with corresponding projection formulas and associated group actions.

We emphasize here that leveraging on hyperbolic projections to cope with Fish-Eye image distortion effects has also been considered in [1] for building deformable convolution kernels through hyperbolic graph embedding techniques. Although anchoring in similar theoretical grounds, the approach we propose here is conceptually different and goes one step further by allowing processing

the Fish-Eye images with equivariant architectures in the 2D hyperbolic space, hence preserving the native geometric dependencies.
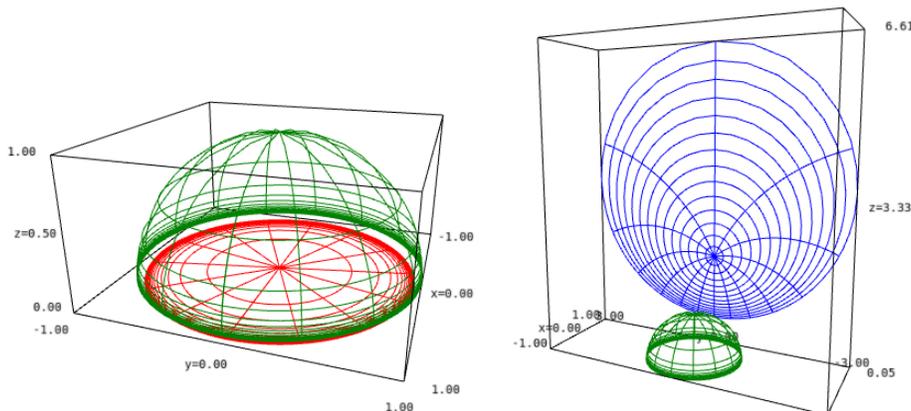
Fig. 2. Link between the Hemisphere model (green), the Poincaré Disk model (red) and the Poincaré Half-Plane model (blue), plotted using the sagemanifolds package `https://sagemanifolds.obspm.fr/index.html`.

2 Hyperbolic Plane Models and Projections

There are several models that are commonly used in practice for representing the 2-dimensional hyperbolic space, including in particular the Hemisphere, the Poincaré Half-Plane and the Poincaré Disk models - see [5] for more details.

The Hemisphere $\mathbb{S}^{\frac{1}{2}}$ model is part of the Riemann Sphere $\mathbb{C}_\infty$ and makes use of its upper half defined by the equation $x^2 + y^2 + z^2 = 1$, for $z > 0$. The Poincaré Disk $\mathbb{D}$ model relies on the open unitary complex disk containing the elements $z = x + iy \in \mathbb{C}$ for which $x^2 + y^2 < 1$. Finally, the Half-Plane model $\mathbb{H}_2$ relies on the upper part of the complex plane consisting in the elements $z = x + iy \in \mathbb{C}$ with $y > 0$. The Poincaré Disk $\mathbb{D}$ can be obtained through the stereographic projection $\pi_{\mathbb{D}}$ of $\mathbb{S}^{\frac{1}{2}}$ from the south pole of $\mathbb{C}_\infty$, i.e. the point of Cartesian coordinates $(0, 0, -1)$, onto the plane $z = 0$. Similarly, the Half-Plane $\mathbb{H}_2$ is obtained through the stereographic projection $\pi_{\mathbb{H}_2}$ of $\mathbb{S}^{\frac{1}{2}}$ from the point $(-1, 0, 0)$ of $\mathbb{C}_\infty$ onto the plane $x = 1$. These 3 different models are represented in Figure 2.

We assume here that the input native Fish-Eye images are given as signals on the hemisphere $\mathbb{S}^{\frac{1}{2}}$, so that a RGB image will be represented by a function $f : \mathbb{S}^{\frac{1}{2}} \to \mathbb{R}^3$, where each component represents one channel. Considering the above formalism, the Fish-Eye image f can be projected onto a signal supported by $\mathbb{D}$ (resp.$\mathbb{H}_2$) by considering $f_{\mathbb{D}} = f \circ \pi_{\mathbb{D}}$ (resp. $f_{\mathbb{H}_2} = f \circ \pi_{\mathbb{H}_2}$). Figures 3

and 4 show an example of the several representations of the same image when considered with support in $\mathbb{S}^{\frac{1}{2}}$, $\mathbb{H}_2$ and $\mathbb{D}$.
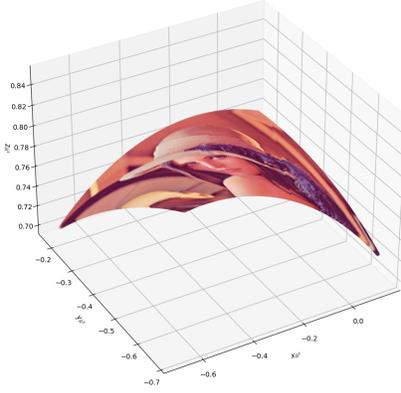
Fig. 3. Lenna picture mapped onto $\mathbb{S}^{\frac{1}{2}}$ by projection into a well chosen tangent plane and then embedded into the 3D Euclidean space $\mathbb{R}^3$

3   Group Action and Equivariance

3.1   Homogeneous Spaces

The Poincaré Half-Plane and Disk are both homogeneous spaces in the sense that they can be written as a quotient space G/H between a given group G and one of its stabilizer subgroup H. More precisely, we have

$$\mathbb{H}_2 = \mathrm{SL}(2,\mathbb{R})/\mathrm{SO}(2) \quad \text{and} \quad \mathbb{D} = \mathrm{SU}(1,1)/\mathrm{U}(1) \tag{1}$$

where we have made use of the following Lie groups, with the matrix multiplication as internal composition law:

$$\mathrm{SU}(1,1) = \left\{ g_{\alpha,\beta} = \begin{bmatrix} \alpha & \beta \\ \bar{\beta} & \bar{\alpha} \end{bmatrix}, \ |\alpha|^2 - |\beta|^2 = 1, \ \alpha,\beta \in \mathbb{C} \right\} \tag{2}$$
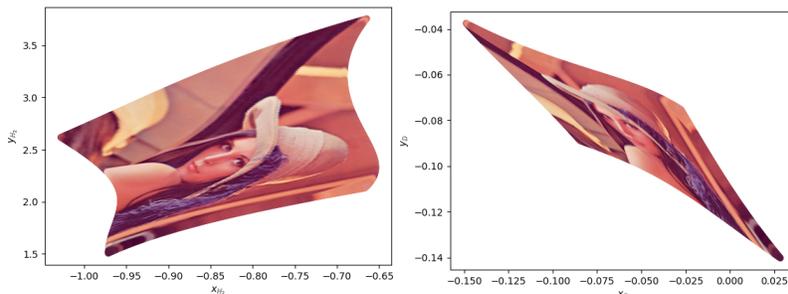
$$\mathrm{SL}(2,\mathbb{R}) = \left\{ g_{a,b} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \ ad - bc = 1, \ a,b,c,d \in \mathbb{R} \right\} \tag{3}$$

$$\mathrm{U}(1) = \left\{ \begin{bmatrix} \frac{\alpha}{|\alpha|} & 0 \\ 0 & \frac{\bar{\alpha}}{|\alpha|} \end{bmatrix}, \ \alpha \in \mathbb{C} \right\} \tag{4}$$

$\mathrm{SU}(1,1)$ (resp. $\mathrm{SL}(2,\mathbb{R})$) has a transitives action $\circ_{\mathbb{D}}$ (resp. $\circ_{\mathbb{H}_2}$) a on $\mathbb{D}$ (resp. $\mathbb{H}_2$) given by:

$$\forall x \in \mathbb{D}, \ g_{\alpha,\beta} \circ_{\mathbb{D}} x = \frac{\alpha x + \beta}{\bar{\beta} x + \bar{\alpha}} \quad \text{and} \quad \forall y \in \mathbb{H}_2, \ g_{a,b} \circ_{\mathbb{H}_2} y = \frac{ay + b}{cy + d} \tag{5}$$

For the action $\circ_{\mathbb{D}}$, U(1) is a stabilizer subgroup fixing the Poincaré Disk center $O_{\mathbb{D}} \in \mathbb{D}$ while SO(2) stabilizes the element $i \in \mathbb{H}_2$ for $\circ_{\mathbb{H}_2}$.

Fig. 4. Projection of the hemispherical Lenna image represented in Figure 3 onto the Poincaré Half-Plane $\mathbb{H}_2$ (left) and the Poincaré Disk $\mathbb{D}$ (right).

3.2   Equivariance and Geometric Deep Learning

Let's first introduce the formal definition of equivariance by considering an operator $F : X \to Y$, where X and Y are two spaces endowed with the actions $\circ_X$ and $\circ_Y$ of a given group G. The operator F is said to be equivariant with respect to the action of G if

$$\forall g \in G, \; \forall x \in X, \; F\left(g \circ_X x\right) = g \circ_Y F\left(x\right) \tag{6}$$

Equivariance is key for Machine Learning applications as highlighted by the success of Convolutional Neural Networks (CNN) for image processing tasks [22], which in particular achieve equivariance with respect to translations through the use of the 2D planar convolution operator. Building Machine Learning architectures achiving more generic equivariance mechanisms and more generally, making use of the native geometry of the inputs data, is an active field of research referred to as Geometric Deep Learning (GDL) [4,12]. More precisely, GDL is getting more and more traction because of its successful application to a wide range of domains [2,11,21]. In this context, Equivariant Neural Networks (ENN) have been shown to be superior to conventional Deep Learning approaches from both accuracy and robustness standpoints and appear as a natural alternative to data augmentation techniques to achieve geometrical robustness with respect to semantically preserving transforms such as isometries. ENN were initially introduced in [6] for image classification by leveraging on group-based equivariant convolution operators and are now achieving state-of-the-art accuracies for a wide range of applications, including for Computer Vision, Graph and Point Cloud processing, Simulation and Trajectory prediction, in Reinforcement

Learning and for Time Series Analysis. Furthermore, ENN are also very appealing from a safety standpoint as achieving robustness-by-design, making them generally promising for Defense related applications [19].

For Fish-Eye image processing, the projection mechanisms onto $\mathbb{D}$ and $\mathbb{H}_2$ and associated group actions previously introduced motivate the use of ENN achieving equivariance with respect to $SU(1,1)$ or $SL(2,\mathbb{R})$. In this context, the approach introduced in [20] can be applied directly for the case of $SU(1,1)$ by projecting the initial Fish-Eye data f to $f_\mathbb{D}$ as mentioned in Section 2 . Leveraging on the formalism of [20], further work will investigate the possibility of building $SL(2,\mathbb{R})$ ENN to handle the projection onto $\mathbb{H}_2$, while other techniques than group-based convolution such as the use of Differential Invariant Theory as in [18] will also be considered in this context.

4   Practical Considerations

In order to project the hemispherical view of a Fish-Eye lens onto the plane, each lens has a projection formula describing the relationship between the incident angle and the position on the projected image. The choice of the distortion being a degree of freedom, several projection formulas exist (see Figure 4 of [3]), with for example $r_d = f\theta$ (equidistant), $r_d = 2f\sin\left(\frac{\theta}{2}\right)$ (equisolid), $r_d = f\sin(\theta)$ (orthographic) and $r_d = 2f\tan\left(\frac{\theta}{2}\right)$ (stereographic). In terms of notations, θ refers here to the angle of incidence measured from the optical axis, f to the focal length and r_d to the distance to the focal plane measured from the optical axis, as shown in the Figure 2 of [16].

We consider in the following that the distortion formula used by the considered lens is specific to it (manufacturer's data or empirical estimate) and that it is given by the following equation:

$$r_d^{f,\theta} = D(f,\theta) \tag{7}$$

where D is a deterministic function such that $\phi : \theta \to D(f,\theta)$ is invertible for a given focal length f. In order to apply the projection techniques described in Section 2, we first need to map the input signal back to the hemisphere through the computation of $\theta = \phi^{-1}\left(r_d^{f,\theta}\right)$.

It should be noted here that in the case where the Fish-Eye lens uses a stereographic distortion, this step only consists in a scaling of the native signal by a factor of $1/(2f)$, in order to bring the projection back into the equatorial plane. These lenses are therefore of particular interest in the context of our approach, although they are not very usual. The Samyang 8 mm f/3.5 and Samyang 8 mm f/2.8 lenses appear to be concrete examples of such lenses[3].

5   Conclusions and Further Work

We have introduced in this paper an approach for Fish-Eye image processing coupling hyperbolic projection techniques and equivariance mechanisms, by in

---

[3] https://en.wikipedia.org/wiki/Fisheye_lens

particular considering the transitive action of $SU(1,1)$ on the Poincaré disk $\mathbb{D}$ and that of $SL(2,\mathbb{R})$ on the Poincaré upper Half-Plane $\mathbb{H}_2$.

Further work will include implementing and evaluating corresponding equivariant neural networks from both accuracy and robustness standpoints, and to conduct some benchmarking with more conventional approaches such as distortion learning through augmentation, transfer learning techniques and the use of deformable kernels.

References

1. Ahmad, O., Lecue, F.: Fisheyehdk:hyperbolic deformable kernel learning for ultra-wide field-of-view image recognition. In: To appear in proceedings of AAAI-22 (2022)
2. Bekkers, E.J., Lafarge, M.W., Veta, M., Eppenhof, K.A., Pluim, J.P., Duits, R.: Roto-translation covariant convolutional networks for medical image analysis (2018)
3. Bettonvil, F.: Fisheye lenses. WGN, Journal of the International Meteor Organization **33**, 9–14 (01 2005)
4. Bronstein, M.M., Bruna, J., Cohen, T., Veličković, P.: Geometric deep learning: Grids, groups, graphs, geodesics, and gauges (2021)
5. Cannon, J.W., Floyd, W., Kenyon, R., Parry, W.R.: Hyperbolic geometry
6. Cohen, T., Welling, M.: Group equivariant convolutional networks. In: Balcan, M.F., Weinberger, K.Q. (eds.) Proceedings of The 33rd International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 48, pp. 2990–2999. PMLR, New York, New York, USA (20–22 Jun 2016), `http://proceedings.mlr.press/v48/cohenc16.html`
7. Cohen, T.S., Geiger, M., Köhler, J., Welling, M.: Spherical cnns. CoRR **abs/1801.10130** (2018), `http://arxiv.org/abs/1801.10130`
8. Coors, B., Condurache, A.P., Geiger, A.: Spherenet: Learning spherical representations for detection and classification in omnidirectional images. In: European Conference on Computer Vision (ECCV) (Sep 2018)
9. Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., Wei, Y.: Deformable convolutional networks (2017)
10. Esteves, C., Allen-Blanchette, C., Makadia, A., Daniilidis, K.: Learning so(3) equivariant representations with spherical cnns. In: Proceedings of the European Conference on Computer Vision (ECCV) (September 2018)
11. Finzi, M., Stanton, S., Izmailov, P., Wilson, A.G.: Generalizing convolutional neural networks for equivariance to lie groups on arbitrary continuous data (2020)
12. Gerken, J.E., Aronsson, J., Carlsson, O., Linander, H., Ohlsson, F., Petersson, C., Persson, D.: Geometric deep learning and equivariant neural networks (2021)
13. Goodarzi, P., Stellmacher, M., Paetzold, M., Hussein, A., Matthes, E.: Optimization of a cnn-based object detector for fisheye cameras. In: 2019 IEEE International Conference on Vehicular Electronics and Safety (ICVES). pp. 1–7 (2019). https://doi.org/10.1109/ICVES.2019.8906325
14. Hawary, F., Maugey, T., Guillemot, C.: Sphere mapping for feature extraction from 360° fish-eye captures. In: MMSP 2020 - 22 nd IEEE International Workshop on Multimedia Signal Processing. pp. 1–6. IEEE, Tempere, Finland (Sep 2020), `https://hal.inria.fr/hal-02924520`

15. He, Y., Yu, H., Yang, W., Scherer, S.: Toward efficient and robust multiple camera visual-inertial odometry (2021)
16. Hughes, C., Denny, P., Jones, E., Glavin, M.: Accuracy of fish-eye lens models. Appl. Opt. **49**(17), 3338–3347 (Jun 2010). https://doi.org/10.1364/AO.49.003338, `http://www.osapublishing.org/ao/abstract.cfm?URI=ao-49-17-3338`
17. Kondor, R., Lin, Z., Trivedi, S.: Clebsch-gordan nets: a fully fourier space spherical convolutional neural network (2018)
18. Lagrave, P.Y., Riou, M.: Toward geometrical robustness with hybrid deep learning and differential invariants theory. Proc. of AAAI 2021 Spring Symposium on Combining Artificial Intelligence and Machine Learning with Physics Sciences, online, March 22-24, 2021 (2021)
19. Lagrave, P.Y., Barbaresco, F.: Introduction to Robust Machine Learning with Geometric Methods for Defense Applications (Jul 2021), `https://hal.archives-ouvertes.fr/hal-03309807`, working paper or preprint
20. Lagrave, P.Y., Cabanes, Y., Barbaresco, F.: "su(1,1) equivariant neural networks and application to robust toeplitz hermitian positive definite matrix classification". In: Nielsen, F., Barbaresco, F. (eds.) Geometric Science of Information. pp. 577–584. Springer International Publishing, Cham (2021)
21. Lagrave, P.Y., Cabanes, Y., Barbaresco, F.: An equivariant neural network with hyperbolic embedding for robust doppler signal classification. In: 2021 21st International Radar Symposium (IRS). pp. 1–9 (2021). https://doi.org/10.23919/IRS51887.2021.9466226
22. Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE **86**(11), 2278–2324 (1998). https://doi.org/10.1109/5.726791
23. Liu, S., Guo, P., Feng, L., Yang, A.: Accurate and robust monocular slam with omnidirectional cameras. Sensors **19**(20) (2019). https://doi.org/10.3390/s19204494, `https://www.mdpi.com/1424-8220/19/20/4494`
24. Meng, L., Hirayama, T., Oyanagi, S.: Underwater-drone with panoramic camera for automatic fish recognition based on deep learning. IEEE Access **6**, 17880–17886 (2018). https://doi.org/10.1109/ACCESS.2018.2820326
25. Playout, C., Ahmad, O., Lecue, F., Cheriet, F.: Adaptable deformable convolutions for semantic segmentation of fisheye images in autonomous driving systems (2021)
26. Rashed, H., Mohamed, E., Sistu, G., Kumar, V.R., Eising, C., El-Sallab, A., Yogamani, S.: Generalized object detection on fisheye cameras for autonomous driving: Dataset, representations and baseline (2020)
27. Su, Y.C., Grauman, K.: Making 360° video watchable in 2d: Learning videography for click free viewing. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1368–1376 (2017). https://doi.org/10.1109/CVPR.2017.150
28. Su, Y.C., Grauman, K.: Kernel transformer networks for compact spherical convolution (2019)
29. Xiao, J., Ehinger, K.A., Oliva, A., Torralba, A.: Recognizing scene viewpoint using panoramic place representation. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. pp. 2695–2702 (2012). https://doi.org/10.1109/CVPR.2012.6247991
30. Yogamani, S., Hughes, C., Horgan, J., Sistu, G., Varley, P., O'Dea, D., Uricar, M., Milz, S., Simon, M., Amende, K., Witt, C., Rashed, H., Chennupati, S., Nayak, S., Mansoor, S., Perroton, X., Perez, P.: Woodscape: A multi-task, multi-camera fisheye dataset for autonomous driving (2021)