Réseaux de neurones hyperboliques équivariants pour le traitement d'images Fish-Eye

Pierre-Yves LAGRAVE[1], Frédéric BARBARESCO[2]

[1]Thales Research and Technology
1 avenue Augustin Fresnel, 91767 Palaiseau, France

[2]Thales Land and Air Systems
9 rue de la Verrerie, 92190 Meudon, France

pierre-yves.lagrave@thalesgroup.com, frederic.barbaresco@thalesgroup.com

Résumé – Le traitement des images Fish-Eye à l'aide d'algorithmes d'apprentissage automatique conventionnels tels que les réseaux de neurones convolutifs est une tâche difficile en raison de la distorsion induite par la projection des images hémisphériques brutes dans le plan euclidien. Nous présentons dans cet article une approche basée sur le domaine émergent de l'apprentissage profond géométrique, dans laquelle des techniques de projection hyperbolique sont couplées à des mécanismes d'équivariance afin de préserver les dépendances géométriques natives et d'obtenir une certaine robustesse par rapport aux variations naturelles de la perception des images Fish-Eye. Ce travail motive en particulier le développement de réseaux de neurones équivariants vis-à-vis des groupes $\mathrm{SU}(1,1)$ et $\mathrm{SL}(2,\mathbb{R})$.

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 $\mathrm{SU}(1,1)$ and $\mathrm{SL}(2,\mathbb{R})$ Equivariant Neural Networks.

1 Introduction et motivations

Les objectifs Fish-Eye sont des capteurs hémisphériques (180 degrés de champ de vision) et la géométrie native des images correspondantes est donc sphérique, conduisant ainsi à des effets de distorsion lors de la projection dans le plan euclidien 2D comme illustré en figure 1. L'utilisation d'algorithmes classiques d'apprentissage profond tels que les réseaux neuronaux convolutionnels (CNN) [20] pour le traitement des images Fish-Eye ne semble donc pas bien adapté et d'autres approches doivent donc être envisagées à cette fin.

Plus précisément, le problème du traitement des images Fish-Eye avec des algorithmes d'apprentissage profond (Deep Learning) suscite de plus en plus d'intérêt car les capteurs Fish-Eye sont utilisés pour des applications pratiques, telles que les tâches de perception pour la conduite autonome [25, 29], l'odométrie visuo-inertielle [15, 21] ou encore le pilotage autonome de drônes [22], les approches considérées pouvant se décomposer selon les 3 familles énoncées ci-dessous :

— Ajustement des algorithmes conventionnels pour prendre en compte les effets de distorsion lors de la phase d'apprentissage effectuée sur les images projetées, y compris l'utilisation de techniques d'augmentation de données [26, 28, 13] et celle d'opérateurs de convolution déformables [8, 24].

— Transfert des algorithmes appris sur des images planaires usuelles à la géométrie sphérique en modifiant à posteriori les caractéristiques algorithmiques (par exemple, les poids, les noyaux de convolution, etc.) [27, 7].

— L'utilisation d'algorithmes opérant directement en géométrie sphérique pour des images omnidirectionnelles (champ de vision de 360°), comme dans [14]. Dans ce contexte, les CNN sphériques [6, 9, 16] atteignent une équivariance par rapport au groupe des rotations 3D $\mathrm{SO}(3)$, les images d'entrée étant représentées comme des signaux sur la sphère 2D. La supériorité de ce type d'approche vis-à-vis des techniques d'augmentation de données a récemment été soulignée dans [12] pour la classification et la segmentation d'images omnidirectionnelles. Bien que fonctionnant dans la géométrie native des entrées, cette approche n'est cependant pas optimale pour les images Fish-Eye avec support hémisphérique.

Bien que certains progrès aient été réalisés afin d'adapter ou de corriger les architectures existantes pour tenir compte des effets de distorsion, aucune méthodologie n'a été proposée pour généraliser la propriété d'équivariance de translation des CNN, qui est hautement bénéfique du point de vue de la précision et de la robustesse lorsqu'on travaille avec des images planaires, au cas des images Fish-Eye, comme cela a été fait avec l'introduction des réseaux neuronaux sphériques pour les images

FIGURE 1 – Première image obtenue en 2021 via les objectifs fish-eye Hazard du rover NASA *Perseverance Mars*. Photo : NASA / JPL-Caltech.

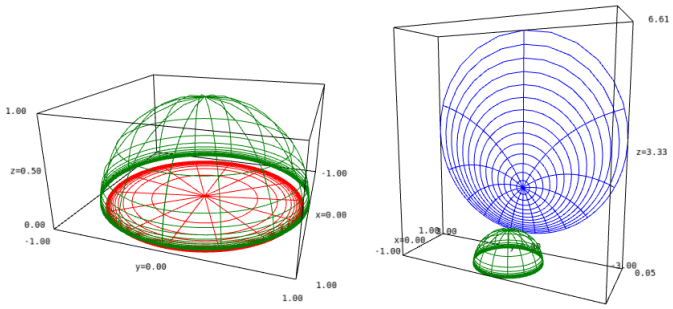

FIGURE 2 – Liens entre le modèle hémisphérique (vert), le modèle du disque de Poincaré (rouge) et le modèle du demi-plan de Poincaré (bleu).

omnidirectionnelles.

Nous proposons donc dans cet article de remédier à cela en utilisant des techniques de projection pour représenter les images natives comme des signaux dans l'espace hyperbolique 2D, puis en construisant des réseaux neuronaux qui sont équivariants [5] vis-à-vis du groupe agissant sur le modèle hyperbolique considéré (par exemple, le disque de Poincaré $\mathbb{D}$, Demi-Plan de Poincaré $\mathbb{H}_2$, etc.), généralisant ainsi l'architecture de CNN sphérique [6]. Nous ne nous attarderons pas ici sur les détails qui sous-tendent la construction de tels réseaux neuronaux (voir par exemple [18] pour des détails concernant les réseaux neuronaux équivariant vis-à-vis de $\mathrm{SU}(1,1)$ pour les signaux définis sur $\mathbb{D}$), mais nous fournissons plutôt un cadre générique pour le traitement des images Fish-Eye en combinant la géométrie hyperbolique avec des mécanismes d'équivariance et en explicitant les formules de projection correspondantes et les actions de groupe associées.

Nous soulignons ici que l'utilisation de la géométrie hyperbolique pour prendre en compte les effets de distorsion pour le traitement d'images Fish-Eye a également été envisagée dans [1] pour construire des noyaux de convolution déformables, en s'appuyant en particulier sur des techniques de plongement de graphes dans le plan hyperbolique. Bien que reposant sur des bases théoriques similaires, l'approche que nous proposons ici est conceptuellement différente et va plus loin en permettant de traiter les images Fish-Eye avec des architectures équivariantes dans l'espace hyperbolique 2D, préservant ainsi les dépendances géométriques natives. Il convient également de noter qu'une approche équivariante aux transformations de Möbius a été très récemment proposée dans [23] pour les images sphériques en considérant l'action de $\mathrm{SL}(2,\mathbb{C})$ sur la sphère Riemannienne $\mathbb{C}_\infty$, confirmant ainsi l'intérêt des approches que nous introduisons ici.

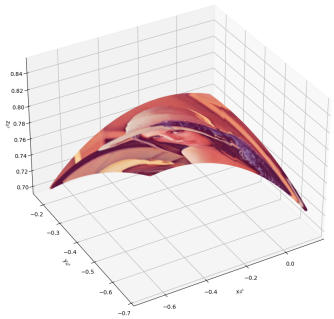

FIGURE 3 – Image de Lenna représentée sur $\mathbb{S}^{\frac{1}{2}}$ et obtenue par projection l'image original dans un plan tangent bien choisi.

2 Modèles du plan hyperbolique et projections

Il existe plusieurs modèles couramment utilisés dans la pratique pour représenter l'espace hyperbolique à 2 dimensions, dont notamment le modèle hémisphérique, le modèle du demi-plan de Poincaré et le modèle du disque de Poincaré - voir [4] pour plus de détails.

Le modèle hémisphérique $\mathbb{S}^{\frac{1}{2}}$ est une partie de la sphère de Riemann $\mathbb{C}_\infty$ et utilise sa moitié supérieure définie par l'équation $x^2+y^2+z^2=1$, pour $z>0$. Le modèle du disque de Poincaré $\mathbb{D}$ s'appuie sur le disque complexe unitaire ouvert contenant les éléments $z=x+iy\in\mathbb{C}$ pour lesquels $x^2+y^2<1$. Enfin, le modèle du demi-plan de Poincaré $\mathbb{H}_2$ s'appuie sur la partie supérieure du plan complexe constituée par les éléments $z=x+iy\in\mathbb{C}$ avec $y>0$. Le Disque de Poincaré $\mathbb{D}$ peut être obtenu par la projection stéréographique $\pi_{\mathbb{D}}$ de $\mathbb{S}^{\frac{1}{2}}$ à partir du pôle sud de $\mathbb{C}_\infty$, c'est-à-dire le point de coordonnées cartésiennes $(0,0,-1)$, sur le plan $z=0$. De même, le demi-plan $\mathbb{H}_2$ est obtenu par la projection stéréographique $\pi_{\mathbb{H}_2}$ de $\mathbb{S}^{\frac{1}{2}}$ du point $(-1,0,0)$ de $\mathbb{C}_\infty$ sur le plan $x=1$. Ces 3 différents modèles sont représentés sur la figure 2.

Nous supposons ici que les images Fish-Eye natives d'entrée sont données comme des signaux définis sur l'hémisphère $\mathbb{S}^{\frac{1}{2}}$, de sorte qu'une image RVB sera représentée par une fonction $f:\mathbb{S}^{\frac{1}{2}}\to\mathbb{R}^3$ où chaque composante représente un canal de couleur. En considérant le formalisme ci-dessus, l'image Fish-Eye f peut être projetée comme un signal défini sur $\mathbb{D}$ (resp. $\mathbb{H}_2$) en considérant $f_{\mathbb{D}}=f\circ\pi_{\mathbb{D}}^{-1}$ (resp. $f_{\mathbb{H}_2}=f\circ\pi_{\mathbb{H}_2}^{-1}$). Les figures 3 et 4 montrent un exemple des différentes représentations de la même image lorsqu'elle est considérée avec un support dans $\mathbb{S}^{\frac{1}{2}}$, $\mathbb{H}_2$ et $\mathbb{D}$.

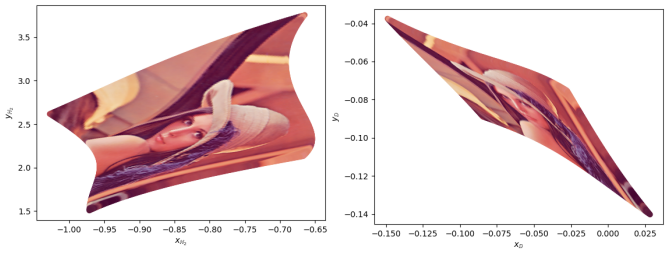

FIGURE 4 – Projection de l'image hémisphérique de Lenna représentée sur la Figure 3 dans le demi-plan de Poincaré $\mathbb{H}_2$ (gauche) et dans le disque de Poincaré $\mathbb{D}$ (droite).

3 Action de groupe et équivariance

3.1 Espaces homogènes

Le demi-plan et le disque de Poincaré sont tous deux des espaces homogènes dans le sens où ils peuvent s'écrire comme un espace quotient G/H entre un groupe donné G et un de ses sous-groupes stabilisateurs H. Plus précisément nous avons $\mathbb{H}_2 = \mathrm{SL}(2,\mathbb{R})/\mathrm{SO}(2)$ et $\mathbb{D} = \mathrm{SU}(1,1)/\mathrm{U}(1)$ où nous avons fait usage des groupes de Lie suivants, avec la multiplication matricielle comme loi de composition interne :

$$\mathrm{SU}(1,1) = \left\{ \begin{bmatrix} \alpha & \beta \\ \bar{\beta} & \bar{\alpha} \end{bmatrix},\ |\alpha|^2 - |\beta|^2 = 1,\ \alpha,\beta \in \mathbb{C} \right\} \quad (1)$$

$$\mathrm{SL}(2,\mathbb{R}) = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix},\ ad - bc = 1,\ a,b,c,d \in \mathbb{R} \right\} \quad (2)$$

$$\mathrm{U}(1) = \left\{ \begin{bmatrix} \frac{\alpha}{|\alpha|} & 0 \\ 0 & \frac{\bar{\alpha}}{|\alpha|} \end{bmatrix},\ \alpha \in \mathbb{C} \right\} \quad (3)$$

Un éléments $g_{\alpha,\beta} \in \mathrm{SU}(1,1)$ agit sur $x \in \mathbb{D}$ via $g_{\alpha,\beta} \circ_{\mathbb{D}} x = \frac{\alpha x + \beta}{\bar{\beta} x + \bar{\alpha}}$. De même, $g_{a,b} \in \mathrm{SL}(2,\mathbb{R})$ agit sur $y \in \mathbb{H}_2$ via $g_{a,b} \circ_{\mathbb{H}_2} y = \frac{ay+b}{cy+d}$. Les deux actions de groupe $\circ_{\mathbb{D}}$ et $\circ_{\mathbb{H}_2}$ sont par ailleurs transitives.

3.2 Apprentissage profond géométrique

Nous commençons tout d'abord par introduire la définition formelle de l'équivariance en considérant un opérateur $F : X \to Y$, où X et Y sont deux espaces munis des actions $\circ_X$ et $\circ_Y$ d'un groupe G donné. L'opérateur F est dit équivariant par rapport à l'action de G si

$$\forall g \in G,\ \forall x \in X,\ F\left(g \circ_X x\right) = g \circ_Y F\left(x\right) \quad (4)$$

La notion d'équivariance est essentielle pour les applications d'apprentissage automatique, comme le montre le succès des réseaux de convolutionnels (CNN) pour les tâches de traitement d'images, car permettant notamment une équivariance par rapport aux translations grâce à l'utilisation de l'opérateur de convolution planaire 2D. La construction d'architectures d'apprentissage automatique offrant des mécanismes d'équivariance plus génériques et, plus généralement, prenant en compte de la géométrie native des données d'entrée, est un domaine de recherche actif appelé *Geometric Deep Learning* (GDL) [3, 11], soit apprentissage profond géométrique en français. Le GDL a été appliqué avec succès dans divers domaines [2, 10, 19]. Dans ce contexte, les réseaux neuronaux équivariants (ENN) [5] se sont révélés supérieurs aux approches classiques d'apprentissage profond et apparaissent comme une alternative naturelle aux techniques d'augmentation des données. Etant par ailleurs robustes par conception, les ENN sont également très intéressants du point de vue sécurité, ce qui les rend généralement prometteurs pour les applications liées à la Défense [17].

Pour le traitement d'images Fish-Eye, les mécanismes de projection sur $\mathbb{D}$ et $\mathbb{H}_2$, ainsi que les actions de groupe associées précédemment introduites motivent l'utilisation d'ENN permettant d'obtenir une équivariance par rapport à $\mathrm{SU}(1,1)$ ou $\mathrm{SL}(2,\mathbb{R})$. Dans ce contexte, l'approche introduite dans [18] peut être appliquée directement pour le cas de $\mathrm{SU}(1,1)$ en projetant les données initiales f en géometrie Fish-Eye vers $f_{\mathbb{D}}$ comme mentionné dans la section 2. Des travaux ultérieurs étudieront la possibilité de construire des ENN équivariants par rapport à $\mathrm{SL}(2,\mathbb{R})$ pour traiter la projection sur $\mathbb{H}_2$, ainsi que l'adéquation entre de telles architectures et les perturbations réelles pouvant altérer les images brutes.

4 Conclusions et travaux futurs

Nous avons introduit dans cet article une approche pour le traitement d'images Fish-Eye couplant des techniques de projection hyperbolique et des mécanismes d'équivariance, en considérant notamment l'action transitive de $\mathrm{SU}(1,1)$ sur le disque de Poincaré $\mathbb{D}$ et celle de $\mathrm{SL}(2,\mathbb{R})$ sur le demi-plan supérieur de Poincaré $\mathbb{H}_2$. Les travaux futurs comprendront l'implémentation et l'évaluation des réseaux de neurones équivariants correspondants, tant du point de vue de la précision que de la robustesse, ainsi qu'une comparaisons avec des approches plus conventionnelles telles que l'apprentissage de la distorsion par augmentation de données, les techniques d'apprentissage par transfert et l'utilisation de noyaux déformables.

Références

[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] Bronstein, M.M., Bruna, J., Cohen, T., Velickovic, P. : Geometric deep learning : Grids, groups, graphs, geodesics, and gauges (2021)

[4] Cannon, J.W., Floyd, W., Kenyon, R., Parry, W.R. : Hyperbolic geometry

[5] 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

[6] Cohen, T.S., Geiger, M., Köhler, J., Welling, M. : Spherical cnns. CoRR **abs/1801.10130** (2018), http://arxiv.org/abs/1801.10130

[7] 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)

[8] Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., Wei, Y. : Deformable convolutional networks (2017)

[9] 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)

[10] Finzi, M., Stanton, S., Izmailov, P., Wilson, A.G. : Generalizing convolutional neural networks for equivariance to lie groups on arbitrary continuous data (2020)

[11] Gerken, J.E., Aronsson, J., Carlsson, O., Linander, H., Ohlsson, F., Petersson, C., Persson, D. : Geometric deep learning and equivariant neural networks (2021)

[12] Gerken, J.E., Carlsson, O., Linander, H., Ohlsson, F., Petersson, C., Persson, D. : Equivariance versus augmentation for spherical images. CoRR **abs/2202.03990** (2022), https://arxiv.org/abs/2202.03990

[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] Kondor, R., Lin, Z., Trivedi, S. : Clebsch-gordan nets : a fully fourier space spherical convolutional neural network (2018)

[17] 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

[18] 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)

[19] 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

[20] 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

[21] 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

[22] 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

[23] Mitchel, T.W., Aigerman, N., Kim, V.G., Kazhdan, M. : Möbius convolutions for spherical cnns. CoRR **abs/2201.12212** (2022), https://arxiv.org/abs/2201.12212

[24] Playout, C., Ahmad, O., Lecue, F., Cheriet, F. : Adaptable deformable convolutions for semantic segmentation of fisheye images in autonomous driving systems (2021)

[25] 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)

[26] Su, Y.C., Grauman, K. : Making 360Â° 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

[27] Su, Y.C., Grauman, K. : Kernel transformer networks for compact spherical convolution (2019)

[28] 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

[29] 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)