

GENERAL E(2) - EQUIVARIANT STEERABLE CNNs

Mohamed Demes et Taher Lmouden

Encadré par: Massinissa Hamidi

ABSTRACT

Avec l'évolution rapide du domaine de l'intelligence artificielle, le besoin de modèles plus puissants et performants ne cesse de croître. Dans le domaine du traitement d'images, la solution la plus couramment utilisée jusqu'à présent est celle des réseaux de neurones convolutifs (CNNs, pour Convolutional Neural Networks). Bien que les CNNs aient démontré leur efficacité en produisant des résultats remarquables dans de nombreuses tâches de traitement d'images, ils présentent néanmoins certaines limitations. Parmi les problèmes bien connus des CNNs classiques figure leur incapacité à gérer efficacement les transformations géométriques, telles que les rotations, les translations, les réflexions, ainsi que leurs combinaisons, dans différents espaces. Ces transformations peuvent considérablement affecter la performance des CNNs, car ces derniers ne reconnaissent pas naturellement qu'un même objet reste identique malgré de telles modifications.

Dans ce contexte, l'approche des CNNs équivariants (Equivariant CNNs) a été proposée pour répondre à ces limitations. Les CNNs équivariants sont conçus pour gérer efficacement ces transformations géométriques et maintenir des performances cohérentes malgré les variations dans les données d'entrée.

Mots clés: Deep Learning, Convolutional Neural Networks, Transformations géométriques, Equivariant CNNs.

Mohamed Demes: 20234747@etud.univ-evry.fr
Taher Lmouden: 20235157@etud.univ-evry.fr

TABLE DE MATIÈRES

1	Introduction	3
1.1	Objectif du Travail d'étude et de recherche	3
2	étude de l'article	3
3	Configuration expérimentale	5
3.1	description du jeux de données	5
3.2	Approches d'entraînement	6
3.2.1	Approche Flattened	6
3.2.2	Approche Par Tâches (Task-based)	6
3.3	Description des modèles	6
3.3.1	CNN Classique	6
3.3.2	CNN Equivariant	7
4	Résultats expérimentaux	7
4.1	Exploration des Modèles CNN Classiques et Équivariants	7
4.2	Impact de la Quantité de Données sur les Performances	8
4.3	Adaptabilité via la Variation des Tasks	9
5	Conclusion	9
	References	10

1 INTRODUCTION

L'un des principaux défis du traitement d'images réside dans la prise en compte des symétries géométriques présentes dans les données réelles, telles que les rotations, les translations et les réflexions. Ces transformations, bien que courantes, représentent une source de difficulté pour les réseaux de neurones convolutifs (CNNs) classiques, qui ne peuvent gérer ces variations qu'en s'appuyant sur des méthodes d'augmentation des données ou un apprentissage explicite pour chaque cas particulier. Cette approche reste coûteuse en termes de temps de calcul et de volume de données requis.

Contrairement aux CNNs classiques, les CNNs équivariants sont capables de capturer et d'exploiter les symétries géométriques des données de manière naturelle. Ces réseaux garantissent que les caractéristiques convolutives se transforment de manière cohérente lorsque l'entrée subit des transformations. En d'autres termes, ils introduisent une notion de correspondance directe entre les transformations de l'entrée et celles des caractéristiques extraites par le réseau. Cela permet d'améliorer les capacités de généralisation des modèles en prenant en compte les variations géométriques des données déjà rencontrées. Cette capacité à intégrer les symétries géométriques de manière explicite confère aux CNNs équivariants des avantages significatifs dans de nombreux domaines. Ils permettent non seulement une meilleure efficacité en termes de données et de calcul, mais également une exploitation plus directe des structures géométriques dans des applications pratiques, telles que la vision par ordinateur, l'imagerie médicale et la robotique.

1.1 OBJECTIF DU TRAVAIL D'ÉTUDE ET DE RECHERCHE

Dans ce travail, nous évaluons les performances des CNNs équivariants en les comparant aux CNNs traditionnels à travers divers défis inspirés de situations réelles. Ces défis incluent des scénarios où les transformations géométriques jouent un rôle critique, afin d'observer leur robustesse et leur capacité à maintenir des performances élevées face à des variations des données d'entrée. Notre objectif est de démontrer dans quelle mesure les CNNs équivariants peuvent surmonter les limitations des CNNs classiques et apporter des solutions plus fiables et efficaces pour des applications pratiques.

2 ÉTUDE DE L'ARTICLE

Les réseaux de neurones convolutifs équivariants (CNNs équivariants) représentent une avancée significative dans le domaine des réseaux de neurones, visant à surmonter les limitations des CNNs classiques. Ces derniers, bien qu'efficaces dans de nombreuses tâches, peinent à gérer naturellement les transformations géométriques des données telles que les rotations, translations ou réflexions. Les CNNs équivariants s'appuient sur des bases mathématiques solides, notamment la théorie des groupes, pour garantir que les caractéristiques extraites se transforment de manière cohérente avec les transformations appliquées à l'entrée, offrant ainsi une meilleure généralisation et une robustesse accrue.

Le concept d'équivariance repose sur le groupe Euclidien $E(2)$, qui modélise les transformations géométriques dans le plan bidimensionnel \mathbb{R}^2 . Ces transformations incluent les rotations continues, les réflexions et les translations, toutes essentielles pour des applications où les orientations des objets varient. Les Steerable CNNs, une des approches les plus avancées dans ce domaine, permettent de concevoir des convolutions équivariantes en imposant des contraintes linéaires sur les kernels. Ces contraintes, dérivées des représentations de groupe, assurent que les champs de caractéristiques respectent les symétries géométriques, un atout majeur dans des contextes où les transformations complexes sont fréquentes.

Parmi les modèles notables, les Group Equivariant CNNs (G-CNNs) ont introduit des symétries discrètes comme les rotations à 90° et les réflexions simples, tandis que les Harmonic Networks ont permis de gérer les rotations continues en utilisant des bases harmoniques. Les $E(2)$ -Steerable CNNs, quant à eux, étendent cette approche à des groupes continus, offrant une plus grande flexibilité pour les transformations géométriques complexes. Une innovation clé est l'introduction de champs scalaires ou vectoriels, appelés champs de caractéristiques, qui décrivent comment les caractéristiques convolutives se transforment sous l'effet des symétries.

Les CNNs équivariants présentent de nombreux avantages dans des applications pratiques. En imagerie médicale, ils permettent de détecter des structures invariantes aux rotations, comme des lésions ou des tumeurs. Dans le domaine de la vision par ordinateur, ils offrent une reconnaissance robuste des objets, indépendamment de leur orientation. En astronomie, ils facilitent l'analyse d'images satellites où les orientations sont imprévisibles. Enfin, dans la robotique, ils s'avèrent précieux pour traiter des données issues de capteurs soumis à des rotations ou translations.

Malgré leurs nombreux avantages, les CNNs équivariants présentent également certaines limites. Leur conception est plus complexe, nécessitant une maîtrise approfondie des concepts mathématiques comme la théorie des groupes. De plus, bien que ces modèles soient puissants pour gérer les transformations géométriques, ils sont généralement optimisés pour des symétries spécifiques et peuvent nécessiter des ajustements pour d'autres types de transformations.

En conclusion, les CNNs équivariants, en particulier les $E(2)$ -Steerable CNNs, constituent une avancée majeure en combinant des bases mathématiques rigoureuses et des applications pratiques. Ils offrent des solutions robustes et efficaces aux défis liés aux transformations géométriques dans des domaines variés, ouvrant ainsi de nouvelles perspectives pour des modèles d'intelligence artificielle plus performants et généralisables.

Un exemple concret pour illustrer les différences entre les CNNs classiques et les CNNs équivariants est présenté ci-dessous. La Figure 1 montre comment un CNN équivariant traite une image soumise à des rotations: les caractéristiques extraites restent cohérentes, indépendamment de l'orientation de l'entrée. En revanche, la Figure 2 montre qu'un CNN classique est fortement affecté par ces rotations, entraînant des variations significatives dans les caractéristiques extraites. Cet exemple met en évidence la capacité des CNNs équivariants à gérer les transformations géométriques de manière naturelle, ce qui les rend particulièrement adaptés pour des applications où les données peuvent subir de telles modifications.

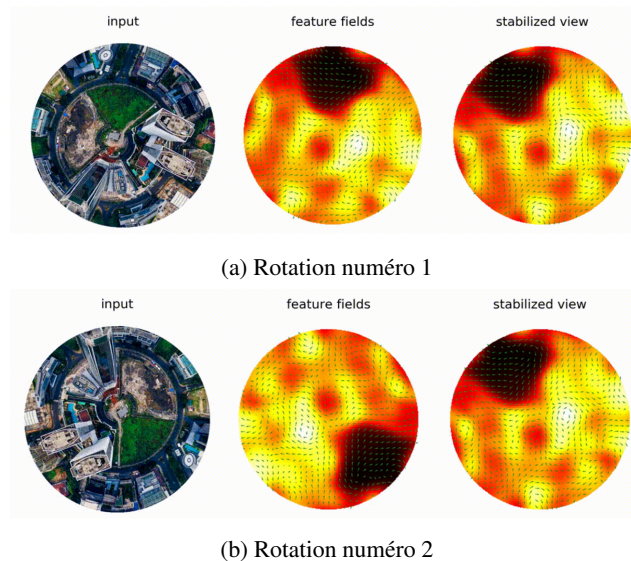


Figure 1: Feature map stable du modèle CNN équivariant pour deux rotations différentes.

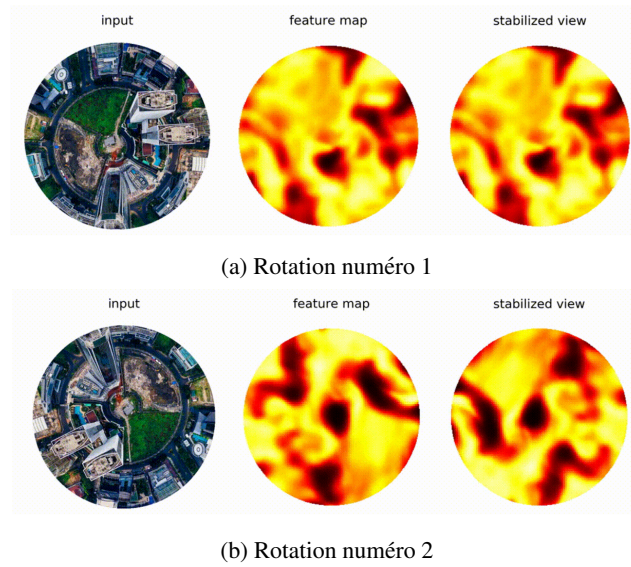


Figure 2: Feature map stable du modèle CNN classique pour deux rotations différentes.

3 CONFIGURATION EXPÉRIMENTALE

3.1 DESCRIPTION DU JEU DE DONNÉES

Le jeu de données utilisé pour les différentes expériences mises en œuvre est le dataset MNIST est un ensemble de données standard utilisé pour les tâches de classification d'images, principalement destiné à la reconnaissance de chiffres manuscrits. Il contient un total de 60000 images pour l'entraînement et 10000 images pour le test, chacune étant une image en niveaux de gris de 28x28 pixels. Les étiquettes (labels) associées aux images représentent les chiffres de 0 à 9, couvrant ainsi 10 classes uniques. Ce dataset est largement utilisé pour évaluer les performances des modèles d'apprentissage automatique, en raison de sa simplicité et de sa pertinence pour les tâches de reconnaissance de formes et de classification.

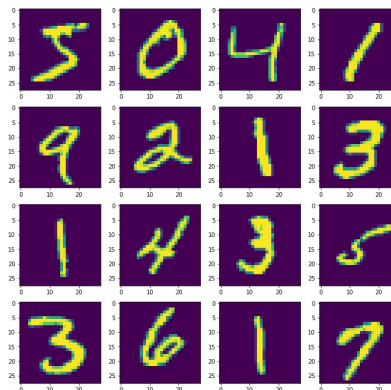


Figure 3: Echantillons d'images de dataset MNIST.

3.2 APPROCHES D'ENTRAÎNEMENT

3.2.1 APPROCHE FLATTENED

Dans cette approche, les images de toutes les classes sont mélangées, et chaque image est associée à une rotation aléatoire sélectionnée parmi un ensemble de rotations définies par le paramètre "pas de rotation" lors du chargement des données. Pour les expériences réalisées, ce pas est fixé à 45° . Le modèle est ainsi entraîné sur un ensemble diversifié d'images présentant des orientations variées, ce qui favorise une généralisation directe à toutes les rotations possibles. La Figure 4 illustre quelques exemples d'images issues de ce dataset enrichi, qui seront utilisées pour entraîner le modèle.

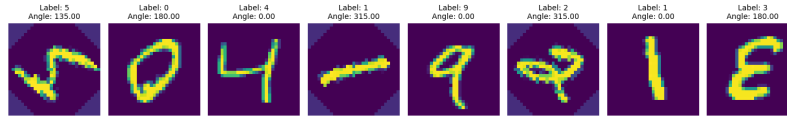
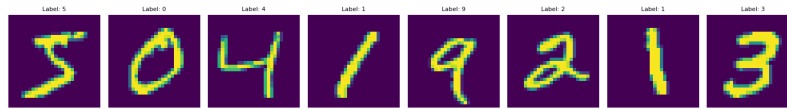


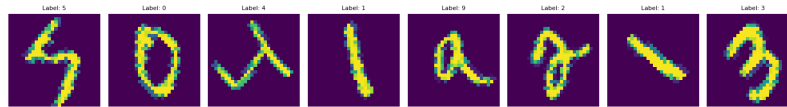
Figure 4: Echantillons d'images de l'approche 'Flattened'.

3.2.2 APPROCHE PAR TÂCHES (TASK-BASED)

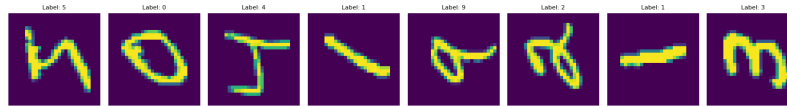
Dans cette approche, le dataset MNIST est utilisé pour créer un ensemble de tâches, où chaque tâche correspond à l'ensemble complet du dataset soumis à une transformation de rotation spécifique. Ces rotations sont définies par un pas fixe de 45° . Le modèle est ensuite entraîné sur l'ensemble de ces tâches, ce qui lui permet d'apprendre à reconnaître les images dans des orientations variées tout en capturant les spécificités propres à chaque rotation. La Figure 5 illustre quelques exemples d'images transformées utilisées lors de l'entraînement.



(a) Task 0, Rotation de 0° .



(b) Task 1, Rotation de 45° .



(c) Task 2, Rotation de 90° .

Figure 5: Echantillons d'images de l'approche 'Task-based'.

3.3 DESCRIPTION DES MODÈLES

3.3.1 CNN CLASSIQUE

Le réseau de neurones convolutif présenté se compose de six blocs de convolution, combinant des couches de convolution, de normalisation et d'activation ReLU. Des couches de pooling sont ajoutées pour réduire la dimensionnalité. Les caractéristiques extraites sont ensuite aplaties et transmises à des couches entièrement connectées pour la classification. Ce modèle a été utilisé pour comparer la robustesse d'un CNN classique face aux transformations à celle des CNN équivariants.

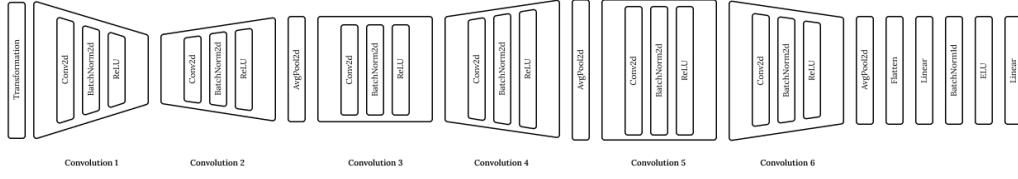


Figure 6: Architecture du CNN Classique utilisé.

3.3.2 CNN EQUIVARIANT

Le CNN équivariant utilisé dans les expériences conserve une structure similaire à celle du CNN traditionnel utilisé. Cependant, contrairement à ce dernier, chaque couche de convolution est spécifiquement conçue pour être équivariante aux transformations. Cela garantit que les sorties du modèle se transforment de manière prévisible lorsque l'entrée subit une modification, comme une rotation, offrant ainsi une meilleure robustesse face aux variations géométriques des données.

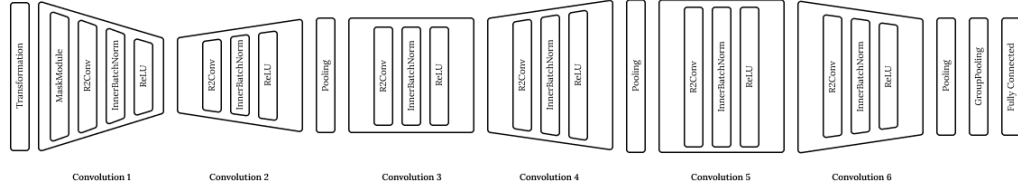


Figure 7: Architecture du CNN Classique utilisé.

4 RÉSULTATS EXPÉRIMENTAUX

Le principal objectif des expériences que nous souhaitons mener est d'explorer et d'évaluer les performances des modèles équivariants. Nous cherchons à analyser comment leurs performances évoluent en fonction des variations des cas d'utilisation, de l'environnement ou des conditions d'entraînement. Cette évaluation vise à mesurer leur robustesse et leur capacité à s'adapter à différents scénarios d'entraînement. Afin de mieux comprendre la puissance des modèles équivariants et d'évaluer leur efficacité, nous réaliserons une comparaison avec les CNNs classiques, ce qui permettra de mettre en évidence les différences et les avantages apportés par les modèles équivariants, notamment en termes de généralisation et de gestion des transformations géométriques.

4.1 EXPLORATION DES MODÈLES CNN CLASSIQUES ET ÉQUIVARIANTS

La première expérience a pour objectif d'explorer et de comparer les performances des modèles CNN classiques et équivariants lorsque l'entraînement est effectué sur l'ensemble complet du dataset. Cette étude vise à analyser si les modèles équivariants surpassent les CNN classiques en termes de performances dans un contexte où les données sont abondantes et disponibles sous plusieurs rotations. L'objectif est de mieux comprendre les comportements des deux types de modèles dans des conditions idéales d'entraînement et d'évaluer l'impact des transformations géométriques sur leurs capacités de généralisation.

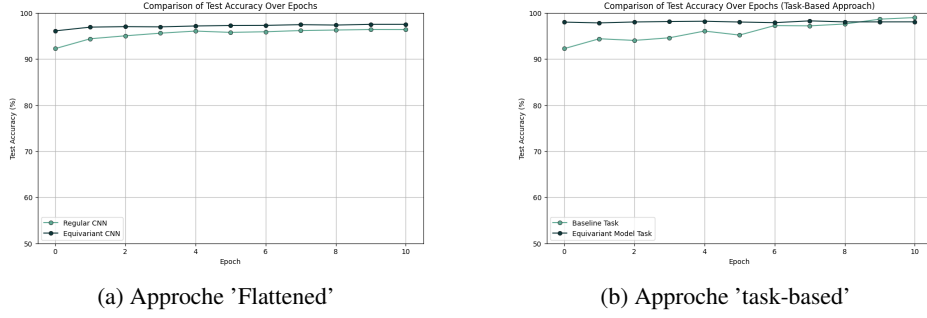


Figure 8: Comparaison des performances des CNNs Classiques et E(2)-CNNs.

La figure 8 montre que les performances des mod les CNN classiques et  quivariants  taient tr s proches, avec des pr dictions pr cises sur l'ensemble de test dans les deux cas. Pour le CNN classique, l'approche "flattened" a permis d'atteindre une pr cision de 96.38%, tandis que l'approche bas e sur les tasks a l g rement am lior  la pr cision   98.97%. De leur c t , les CNNs  quivariants ont obtenu des r sultats comparables, avec une pr cision de 97.51% pour l'approche "flattened" et de 98.05% pour l'approche task-based. Ces r sultats confirment que les deux mod les sont capables de g n raliser efficacement dans ce contexte.

4.2 IMPACT DE LA QUANTIT  DE DONN ES SUR LES PERFORMANCES

Dans cette deuxi me exp rience, nous analysons comment la quantit  de donn es d'entra nement influence les performances des CNN classiques et  quivariants. Les mod les seront entra n s sur des sous-ensembles du dataset flattened, avec des quantit s de donn es allant de 0.5%   100%.   chaque  tape, la quantit  de donn es sera progressivement augment e (0.5%, 10%, 20%, 30%, etc.), et les performances seront mesur es pour chaque niveau. Cette analyse permettra de comparer la robustesse des deux types de mod les face   des r ductions significatives des donn es d'entra nement et d' valuer leur capacit  de g n ralisation dans des conditions de donn es limit es.

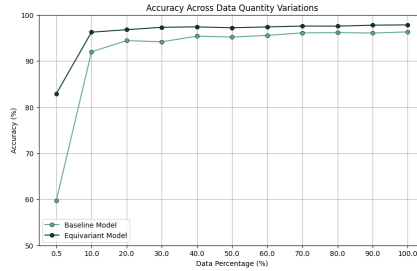


Figure 9: Accuracy en fonction de quantit  de donn es.

D'apr s la figure X, on observe que, d s que 10% du dataset d'entra nement est utilis , les performances des mod les CNN  quivariants et CNN classiques sont proches et ne pr sentent pas de diff rences significatives. Cela s'explique par la grande taille du dataset MNIST, qui contient 60,000 images: 10% de cet ensemble repr sente 6,000 images, un nombre cons quent pour l'entra nement, permettant aux deux mod les de bien g n raliser gr ce   la diversit  des donn es rencontr es.

Pour aller plus loin, nous avons r duit de mani re drastique le pourcentage de donn es d'entra nement, en utilisant seulement 0.5% du dataset, soit un total de 300 images,  quivalant   30 images par classe. Ce sc nario teste r ellement la capacit  de g n ralisation de chaque type de mod le. Dans ce cas, une diff rence marqu e est apparue entre les deux mod les: le E(2)-CNN a atteint une pr cision de 82.9%, tandis que le CNN classique n'a pas d pass  59.75%.

Table 1: Conclusion: accuracy des deux modèles en fonction de la quantité de données.

Quantité de données	0.5%	10%	20%	40%	50%	70%	90%	100%
CNN Classique	59.75%	91.98%	94.42%	95.38%	95.18%	96.09%	96.05%	96.30%
CNN Equivariant	82.9%	96.24%	96.78%	97.39%	97.20%	97.57%	97.76%	97.82%

4.3 ADAPTABILITÉ VIA LA VARIATION DES TASKS

La troisième expérience se concentre sur l'évaluation de l'adaptabilité des modèles CNN équivariants et classiques en fonction d'un nombre croissant de rotations, appelées "tasks". Initialement, le modèle est entraîné sur une seule task, correspondant à une seule rotation spécifique, puis testé sur l'ensemble des rotations disponibles (huit dans notre cas). Par la suite, l'entraînement est effectué en augmentant progressivement le nombre de tasks: deux rotations, trois rotations, et ainsi de suite, jusqu'à couvrir l'ensemble des huit rotations. L'objectif est de mesurer dans quelle mesure les modèles, en particulier les CNN équivariants, peuvent généraliser et s'adapter à des rotations non vues en utilisant le minimum de nombre de rotations disponibles.

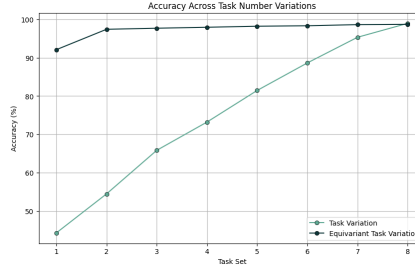


Figure 10: Accuracy en fonction de quantité de données.

D'après la figure X, on constate que le modèle E(2)-CNN, avec un apprentissage basé sur une seule rotation, a réussi à généraliser efficacement sur l'ensemble de test comprenant des images avec toutes les rotations. En effet, avec une seule rotation, il a atteint une précision de 92.1%, tandis que le CNN classique a montré des performances nettement inférieures, n'excédant pas 44.27%.

En utilisant deux rotations (ou tâches), le modèle équivariant a pu atteindre des performances comparables à celles obtenues avec l'ensemble complet des tâches, démontrant sa capacité à généraliser à partir d'un petit sous-ensemble de rotations. En revanche, le modèle CNN classique nécessite l'ensemble complet des rotations pour maintenir des performances acceptables, ce qui souligne sa dépendance à une couverture exhaustive des transformations pendant l'entraînement.

5 CONCLUSION

Les CNNs équivariants reposent sur une base théorique solide qui les distingue fondamentalement des CNNs classiques. En s'appuyant sur la théorie des groupes, ils exploitent les symétries géométriques présentes dans les données pour garantir que les caractéristiques convolutives se transforment de manière cohérente face aux rotations, réflexions, et translations. Cette propriété, au cœur des Steerable CNNs, est obtenue en imposant des contraintes spécifiques sur les kernels, assurant ainsi une généralisation efficace aux transformations géométriques. Comme exploré dans la section 2, ces approches théoriques permettent de surmonter les limitations structurelles des CNNs classiques face aux variations complexes des données d'entrée.

Sur le plan pratique, les expériences réalisées ont montré que les CNNs équivariants offrent des avantages significatifs, notamment dans des contextes où les données sont limitées. Ils démontrent une robustesse et une capacité de généralisation nettement supérieures aux CNNs classiques, même avec un faible nombre d'exemples d'entraînement. De plus, leur adaptabilité leur permet de gérer

efficacement des scénarios où toutes les transformations ne sont pas vues pendant l'entraînement, montrant ainsi leur efficacité dans des environnements contraints.

En résumé, les CNNs équivariants offrent une solution robuste pour des tâches nécessitant une gestion explicite des symétries géométriques et une généralisation à partir de données limitées. Leur potentiel les rend particulièrement adaptés à des domaines tels que l'imagerie médicale, la vision par ordinateur et la robotique, où les transformations géométriques jouent un rôle crucial.

REFERENCES

Maurice Weiler and Gabriele Cesa. General $e(2)$ -equivariant steerable cnns. *CoRR*, abs/1911.08251, 2019. URL <http://arxiv.org/abs/1911.08251>.