

Reconstruire le passé : Identifier des paires de fragments adjacents à l'aide de l'apprentissage profond

Étienne Beaulac, Fadel Touré et Alain Goupil

Université du Québec à Trois-Rivières – Département de mathématiques et informatique

Introduction

Lors de leurs recherches, les archéologues découvrent de nombreux artéfacts, souvent brisés en plusieurs morceaux. Au terme de leurs fouilles, ils se retrouvent donc, en quelque sorte, avec plusieurs casse-têtes dont les pièces ont été mélangées ensemble.

La littérature présente plusieurs approches utilisant l'apprentissage profond pour reconstruire des objets à partir de leurs fragments. Or, plusieurs de ces approches restreignent le nombre, la forme ou l'orientation des fragments, ce qui ne correspond pas à la réalité des archéologues.

L'objectif de ce projet de recherche est donc de remédier à ces limitations en développant une approche basée sur l'apprentissage profond et en mesure de comparer deux fragments de forme et d'orientation quelconques afin de déterminer s'ils sont adjacents.

Données

Comme il n'existe pas de grand jeu de données public qui réponde à nos besoins, les données qui serviront à l'entraînement des réseaux de neurones ont été générées artificiellement. Des patrons de fracture sont générés à partir de bruits aléatoires, puis appliqués successivement à l'image d'origine pour la briser en de multiples morceaux (figure 1).



Fig. 1 – Quatre fragments générés artificiellement par deux patrons aléatoires de fracture. Ils forment ensemble quatre paires adjacentes.

À l'aide de cet algorithme, 950 carrés de couleur unie ont été fracturés en 4 morceaux. 3 800 fragments ont été générés, ce qui forme également 3 800 paires adjacentes. Celles-ci ont été séparées en deux sous-ensembles : 3 200 pour l'entraînement et 600 pour la validation.

Méthodologie

L'objectif du projet nécessite la comparaison entre deux images. L'architecture de notre réseau de neurones sera donc **siamoise** : elle contiendra deux branches jumelles, une pour chaque fragment de la paire. Ces deux branches sont ensuite fusionnées pour produire la sortie finale du réseau : la probabilité entre 0 et 100 % que les deux fragments reçus soient adjacents.

Notre réseau contient des couches convolutives : elles permettent d'extraire des caractéristiques d'images. Ces couches sont équivariantes aux translations, c'est-à-dire que le réseau pourra extraire les mêmes caractéristiques d'une image ayant subi un déplacement horizontal ou vertical. Cependant, la convolution n'est pas équivariante aux rotations (figure 2). Cela pose problème pour les fragments de notre jeu de données, dont l'orientation peut varier.

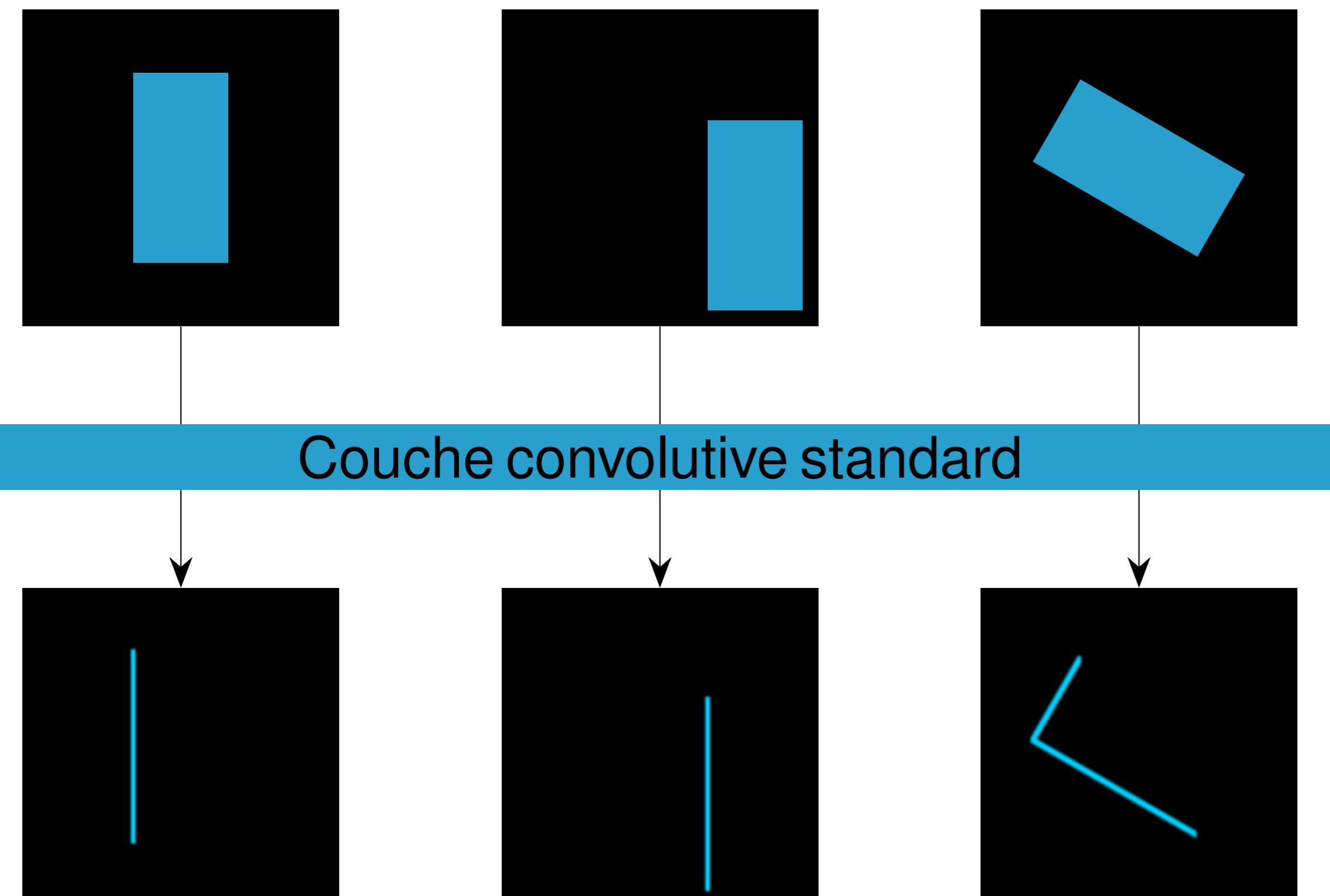


Fig. 2 – À gauche : une couche convulsive extrait les côtés gauches des formes. Au centre : lorsque l'image en entrée subit une translation, les caractéristiques extraites subissent la même translation (équivariance). À droite : une convolution n'est pas équivariante aux rotations.

Pour résoudre ce problème, notre réseau utilise plutôt un nouveau type de convolution qui est équivariant aux rotations [1, 2]. En choisissant judicieusement le nombre de ces couches, leur taille et les opérations mathématiques entre elles, il est possible de rendre notre réseau siamois **invariant aux rotations**. Ainsi, avant même de débuter l'entraînement, nous avons la garantie que la sortie du réseau sera la même, peu importe l'orientation des fragments.

Méthodologie (suite)

Pour simplifier le travail du réseau, les images des fragments ne sont pas envoyées en entier. Ce sont plutôt des sous-images des fragments, prises sur leur frontière, qui sont comparées.

Une fois que le réseau est entraîné, on peut l'utiliser pour comparer deux fragments dans leur entièreté. Il faut alors prendre toutes les sous-images prises le long de la frontière du premier fragment, puis les comparer à toutes les sous-images du second. En stockant les probabilités obtenues dans un tableau, on génère la **matrice d'adjacence** de la paire. La matrice d'adjacence d'une paire adjacente comportera une diagonale (figure 3) avec des probabilités élevées, ce qui n'est pas le cas pour deux fragments non adjacents (figure 4).

Ces matrices d'adjacence forment alors un nouveau jeu de données pouvant servir à l'entraînement d'un second réseau de neurones. C'est lui qui prendra la décision finale quant à l'adjacence d'une paire de fragments en se fiant à leur matrice d'adjacence générée par le premier réseau.

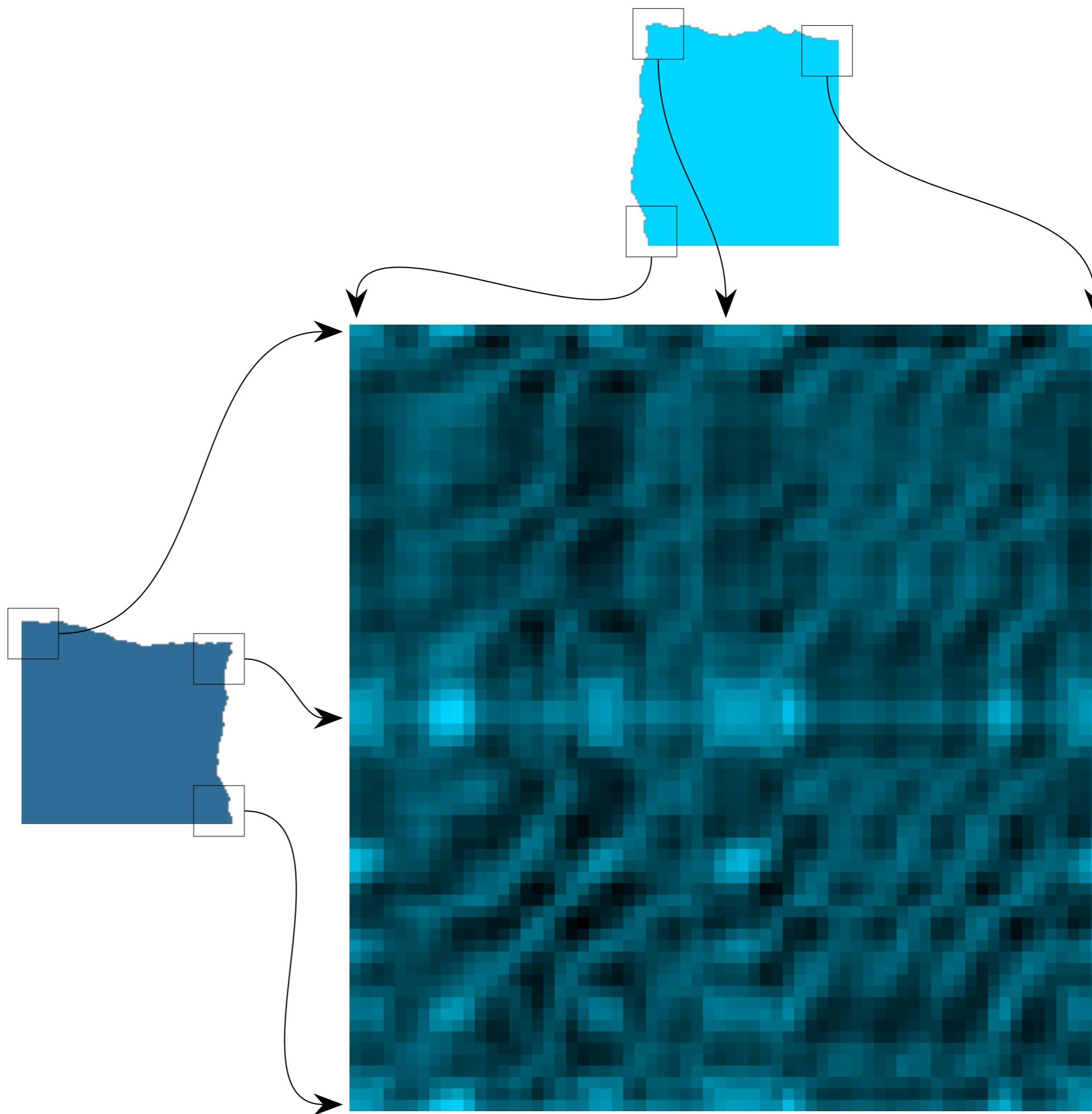


Fig. 3 – Matrice d'adjacence de deux fragments adjacents. Le quadrant inférieur gauche contient une diagonale parfaite.

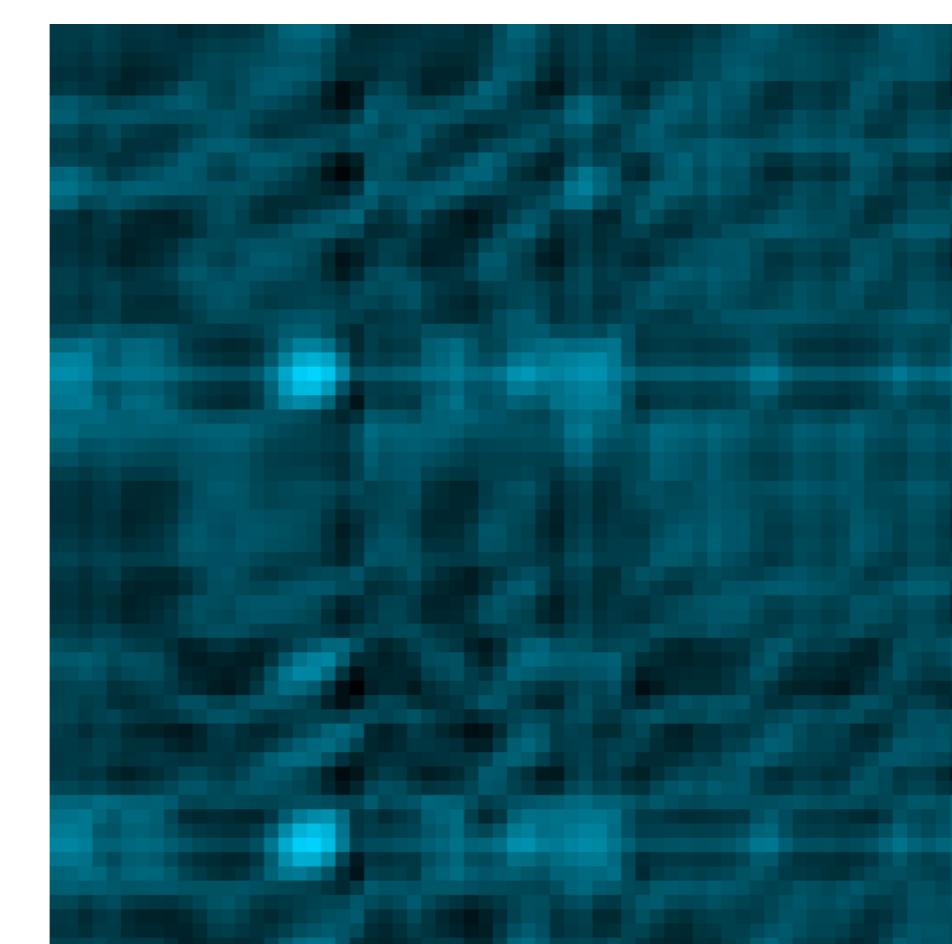


Fig. 4 – Matrice d'adjacence de deux fragments non adjacents. Il n'y a pas présence de diagonale parfaite.

Résultats

Les entraînements ont montré qu'une taille de sous-image plus petite (ici, 33 pixels de côté) rend l'apprentissage plus rapide et augmente la précision. Sur les données de validation, le premier réseau (adjacence de sous-images) atteint une précision de **84,75 %**, tandis que le deuxième réseau (classification des matrices d'adjacence) atteint une précision de **98,33 %**.

Conclusion

Il est possible d'utiliser l'apprentissage profond pour déterminer si des paires de fragments 2D sont adjacentes, et ce, sans restreindre la forme et l'orientation des fragments. Notre méthodologie, utilisant deux réseaux de neurones, a démontré son efficacité avec sa précision de 98,33 %. Dans de futurs travaux, il serait intéressant de généraliser l'approche pour les fragments 3D en tenant compte de leur couleur et, en particulier, de leur épaisseur.

Références

- [1] T. S. Cohen et M. Welling, "Group equivariant convolutional networks", dans *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, 2016, p. 2990-2999.
- [2] M. Weiler et G. Cesa, "General $E(2)$ -Equivariant Steerable CNNs", dans *Conference on Neural Information Processing Systems (NeurIPS)*, 2019.

Remerciements