Agrégation par analogie : reconstruction et classification d'image

Jakub Pilimon[1], Daiyu Ma[1], Miguel Couceiro[2], Yves Lepage[1]

[1] Université Waseda, Japon
[2] INESC-ID, IST, Université de Lisbonne, Portugal

pilimon.jakub@akane.waseda.jp, madayu@asagi.waseda.jp,
miguel.j.couceiro@tecnico.ulisboa.pt, yves.lepage@waseda.jp

Résumé

Cet article explore les relations entre l'analogie numérique et le processus d'abstraction au travers d'applications en traitement d'image. Il introduit une fonction d'agrégation par analogie qui réduit les images mais produit aussi des informations spécifiques à l'analogie numérique. Il montre comment recombiner ces informations pour la reconstruction d'image et examine les effets de l'agrégation-reconstruction sur la classification. Cette étude révèle de façon expérimentale le fort potentiel de l'analogie numérique pour la généralisation.

Mots-clés

Analogie, traitement de l'image, agrégation, reconstruction, classification.

Abstract

This article explores the relationship between numerical analogy and abstraction through applications in image processing. It introduces an analogy-based pooling function that reduces images but also produces information specific to numerical analogy. It shows how to recombine all information in image reconstruction and examines the effects of pooling-reconstruction on classification. This study experimentally reveals the strong potential of digital analogy for generalization.

Keywords

Analogy, image processing, pooling, reconstruction, classification.

1 Introduction

La généralisation est le principal objectif de l'apprentissage automatique. Les méthodes d'apprentissage automatique reposent sur l'hypothèse que les données d'entraînement et de test sont tirées de la même distribution de probabilité. Quand la distribution des données de test dévie par trop de la distribution d'entraînement, les performances des modèles se dégradent [16, 13].

Dans les tâches de reconnaissance d'image, les décalages de distribution peuvent résulter de changements dans la qualité de l'image, de bruit ou d'autres variations qui modifient les caractéristiques d'entrée sans affecter la sémantique de classe sous-jacente. En conséquence, les modèles qui fonctionnent bien sur la distribution d'apprentissage peuvent échouer à généraliser lorsque des caractéristiques visuelles telles que la netteté ou certains détails changent.

On considère souvent que l'abstraction est au cœur de la généralisation. C'est donc une faculté importante à modéliser en intelligence artificielle. L'abstraction peut être comprise comme un processus par lequel l'information s'exprime sous une forme différente, mais permet néanmoins la reconstruction d'objets similaires à ceux de départ. Dans certains test d'intelligence, par exemple des puzzles à résoudre, elle est au cœur de la création des solutions.

Le raisonnement par analogie est la capacité à identifier des ressemblances ou des différences récurrentes entre des choses, des situations ou des éléments constitutifs d'un problème, à les analyser et à en déduire des généralisations ou des conclusions. Le raisonnement par analogie étant considéré comme un indicateur d'intelligence, les analogies sont utilisées dans les tests de mesure de quotient intellectuel dont les matrices progressives de Raven [11] constituent un exemple. La difficulté à résoudre de tels tests réside dans le fait qu'un raisonnement abstrait de recombinaison par analogie est indispensable à la production de la solution.

Cet article se propose d'explorer l'abstraction dans la réduction et la reconstruction d'image à l'aide d'une opération d'agrégation fondée précisément sur l'analogie et de tester la qualité de cette reconstruction au moyen de la classification. Cette opération d'agrégation peut être vue comme un processus d'abstraction puisqu'elle permet de séparer les informations contenues dans un objet en deux types d'informations, le premier correspondant à un objet similaire réduit, le second correspondant à des informations spécifiques à l'analogie numérique. Recombinées à l'objet réduit, elles devraient permettre la reconstruction de l'objet de départ. L'objet reconstruit peut ne pas être exactement identique à l'objet de départ, car on espère qu'un peu de généralisation s'y glisse grâce à l'abstraction qui en a été faite, d'où un décalage éventuel de distribution.

Les contributions de cet article consistent en l'exposé de la méthode d'agrégation par analogie et dans sa mise à l'épreuve dans la reconstruction d'images dont on testera la déviation par comparaison directe avec les objets de départ mais aussi le degré de généralisation par la performance en classification de modèles construits à partir d'objets reconstruits ou pas sur des objets reconstruits ou pas.

2 L'analogie numérique

L'utilisation de l'analogie a longtemps été limitée aux configurations booléennes ou réduite à l'analogie arithmétique. L'analogie booléenne restreint l'analogie à des configurations discrètes, c'est-à-dire non continues. L'analogie arithmétique, qui s'applique au continu, limite elle l'interprétation à une égalité des différences et suppose la linéarité de l'espace de représentation. La notion de moyenne généralisée lève les limitations précédentes [8, 9, 3]. Pour tout $p \in \mathbb{R} \cup \{-\infty, +\infty\}$, la moyenne généralisée $m_p(x_1, \ldots, x_n)$ de n nombres réels positifs ou nuls $x_1, \ldots x_n$ est définie comme suit.

$$m_p(x_1, \ldots, x_n) = \lim_{r \to p} \left(\frac{1}{n} \sum_{i=1}^{n} {x_i}^r \right)^{1/r} \tag{1}$$

La moyenne généralisée correspond à la moyenne arithmétique pour $p = 1$; à la moyenne harmonique pour $p = -1$; à la moyenne quadratique pour $p = 2$; à la moyenne géométrique lorsque p tend vers 0 [1] ; au maximum ou au minimum lorsque p tend vers $+\infty$ ou $-\infty$. Les moyennes généralisées sont des moyennes quasi-arithmétiques [4] et possèdent donc quelques bonnes propriétés.

On dira qu'il existe une *analogie numérique* [8] en puissance p entre quatre nombre réels a, b, c et d de $\mathbb{R}_+$, et l'on écrira $a : b ::^p c : d$, s'il existe un p dans $\mathbb{R} \cup \{-\infty, +\infty\}$ tel que

$$m_p(a, d) = m_p(b, c). \tag{2}$$

Selon la terminologie classique, les termes a et d d'une analogie sont appelés les *extrêmes* et les termes b et c les *moyens*. On appellera p la *puissance* de l'analogie numérique.

Moyenne analogique. On posera que la moyenne analogique de quatre nombres réels positifs ou nuls $0 \leq a \leq b \leq c \leq d$ est la moyenne généralisée $m_p(a, b, c, d)$ de ces quatre nombres. La définition de l'analogie numérique et les propriétés des moyennes quasi-arithmétiques produisent facilement que $m_p(a, b, c, d) = m_p(a, d) = m_p(b, c)$. Par exemple, comme il existe une analogie $2 : 4 ::^p 6 : 8$ avec $p = 1$ entre les valeurs $2, 4, 6, 8$, leur moyenne analogique est $m_1(2, 4, 6, 8) = m_1(2, 8) = m_1(4, 6) = 5$.

Ordonnancement. La symétrie des arguments dans les moyennes généralisées, propriété des moyennes quasi-arithmétiques, permet d'échanger les deux extrêmes ou les deux moyens dans la définition (2). Combinée avec la symétrie de l'égalité, il en résulte huit formes qui correspondent aux huit formes équivalentes d'une même analogie et les 24 combinaisons possibles pour quatre termes donnés se réduisent à trois ordonnancements distincts possibles $a : b ::^p c : d$, $a : c ::^p d : b$ et $a : b ::^p d : c$ pouvant chacun conduire à une analogie (voir [8]).

Puissance analogique. Il a été dit plus haut que, dans une analogie numérique, p est la puissance de l'analogie. Pour des termes positifs et différents deux à deux, on peut montrer qu'un tel p existe et est unique à condition de trier les termes par ordre croissant (de nouveau, voir [8]). Mentionnons cependant un cas très particulier, celui où tous les termes de l'analogie sont égaux. Dans ce cas, les trois ordonnancements sont tous possibles et tout p de $\mathbb{R} \cup \{-\infty, +\infty\}$ permet d'écrire $a : a ::^p a : a$.

1. $\lim_{p \to 0} m_p(x_1, \ldots x_N) = \sqrt[N]{\prod_{i=1}^{N} x_i}$.

3 Agrégation

L'agrégation est présente dans les réseaux de neurones convolutifs appliqués au traitement d'image depuis le début [6, 5]. Elle est à la base de l'architecture U-Net [12]. Elle consiste à remplacer un bloc de pixels par la valeur d'une fonction calculée sur les valeurs des pixels du bloc. Comme le bloc est remplacé par un pixel contenant cette valeur, l'agrégation réduit la taille de l'image.

Dans les agrégations par maximum ou par moyenne, un bloc de pixels est simplement remplacé par le maximum ou la moyenne, pris sur ce bloc.

L'agrégation par moyenne généralisée est une généralisation des deux méthodes précédentes [7, 17]. Elle repose sur l'observation que la valeur maximale est la moyenne généralisée pour $p = +\infty$ et que la moyenne arithmétique est la moyenne pour $p = 1$. Son introduction dans les réseaux de neurones convolutifs en traitement d'image repose sur l'idée qu'un réseau de neurones se devrait de déterminer lui-même quelle moyenne généralisée appliquer, c'est-à-dire, quel p de $\mathbb{R} \cup \{-\infty, +\infty\}$ choisir. En conséquence, pour chaque bloc, une valeur différente de p est déterminée par rétro-propagation.

Agrégation par analogie. Nous proposons l'agrégation par analogie qui, comme les agrégations par moyenne ou par moyenne généralisée, remplace un bloc de pixels par un seul pixel, avec effet de réduire la taille de l'image. Mais c'est une variante, à deux égards, de l'agrégation par moyenne généralisée. Premièrement, nous supprimons la détermination de la puissance appropriée par apprentissage, et imposons sa détermination par l'analogie numérique existant entre les valeurs du bloc. Il n'y a ici aucune rétro-propagation. Deuxièmement, l'utilisation de l'analogie implique une limitation de la taille des blocs à quatre pixels, car une analogie ne comporte que quatre termes.

Produit de l'agrégation par analogie. Pour une image carrée de $n \times n$ pixels, [2] l'agrégation par analogie produit deux types d'information. Le premier type est une image réduite de $(n-1) \times (n-1)$ pixels où chaque pixel vient en remplacement d'un bloc de 2×2 pixels de l'image de départ et prend la valeur de la moyenne analogique des pixels de ce bloc. Le deuxième type d'information comprend l'ordonnancement et la puissance de l'analogie correspondant à chaque bloc de l'image de départ. Ces deux types d'information sont illustrés dans la figure 1. En pratique, pour chaque bloc carré de 2×2 pixels de l'image de départ, la configuration des extrêmes et des moyens dans le bloc suit l'une de trois configurations possibles, que l'on fait corres-

2. Nous nous restreignons ici aux images carrées pour la simplicité de l'exposé, mais la généralisation à des images de hauteur et largeur différentes, et possédant même plusieurs canaux, est aisée.

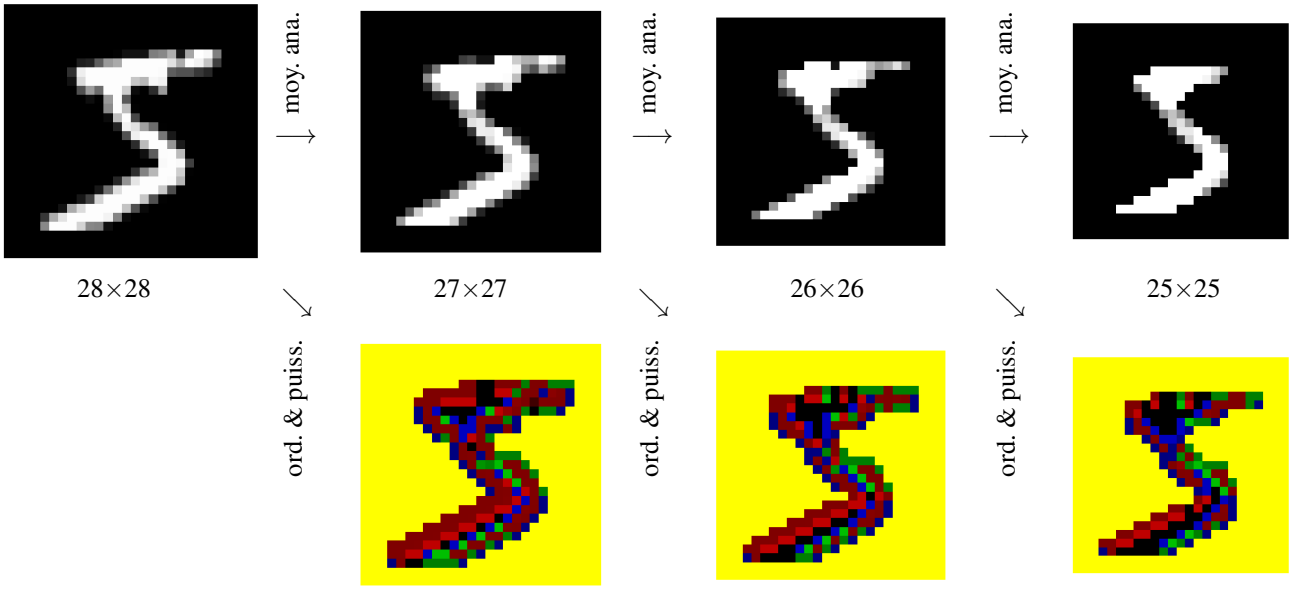

FIGURE 1 – Agrégation par analogie appliquée itérativement 3 fois. L'image de départ en haut à gauche comporte 28×28 pixels et les images suivantes sont de taille 27×27, 26×26 et 25×25 pixels respectivement car à trois profondeurs d'agrégation successives. Les pixels dans images du haut ont pour valeurs les moyennes analogiques des blocs de 2×2 pixels de l'image précédente. Les images du bas visualisent les ordonnancements et les puissances. Les couleurs *vert, bleu* et *rouge* distinguent les trois ordonnancements possibles (voir figure 2), la couleur étant plus foncée pour une puissance plus élevée. Les pixels de couleur *jaune* proviennent de blocs de pixels de valeur égale.

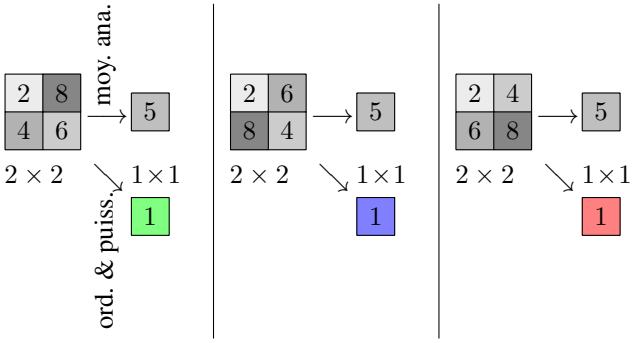

FIGURE 2 – Trois exemples de configurations possibles pour les valeurs des extrêmes (2 et 8) et des moyens (4 et 6) dans un bloc carré de 2×2 pixels. L'analogie est arithmétique car $\frac{1}{2}(2+8) = \frac{1}{2}(4+6)$, donc $p = 1$, et la moyenne est $m_1(2,4,6,8) = m_1(2,8) = m_1(4,6) = 5$. Mais de gauche à droite, la disposition des extrêmes (2 et 8) est horizontale, verticale ou en diagonale, d'où des couleurs différentes pour indiquer des ordonnancements différents.

pondre aux trois ordonnancements (voir figure 2). La puissance de l'analogie peut alors être déterminée, ainsi que la moyenne analogique.

Profondeur d'agrégation. L'agrégation par analogie peut être appliquée de manière itérative à une image carrée jusqu'à réduire l'image à un seul pixel. Ce processus peut être arrêté à n'importe quelle *profondeur*. La figure 1 illustre l'application itérative de l'agrégation par analogie à une image grisée à un seul canal, jusqu'à une profondeur de trois.

Agrégation par analogie et abstraction. L'agrégation réduit et transforme une image en lui donnant une autre structure. Ici, elle éclate l'image en deux autres images différentes mais ces deux images peuvent aussi être interprétées comme une seule image à quatre canaux. L'agrégation par analogie, si elle réduit certes la taille de l'image, ajoute de l'information puisque la quantité d'information livrée par les quatre canaux est supérieure à celle livrée par l'image de départ. Il ne s'agit pas d'une compression.

4 Reconstruction d'image

L'agrégation par analogie, en tant que processus d'extraction et de restructuration de l'information, peut être vue comme un processus d'abstraction. Nous nous intéressons à savoir si elle induit une généralisation quelconque. Afin d'examiner cette hypothèse, nous examinons la possibilité de reconstruire l'image de départ à partir des informations extraites et restructurées à chaque profondeur d'agrégation.

La qualité de reconstruction d'une image se mesure selon des critères qui se réduisent tous à savoir en quoi l'image reconstruite est proche de l'image d'origine. Après avoir utilisés de tels critères qui se contentent de comparer l'image reconstruite à l'image de départ, on se demandera ensuite si la variabilité observée est laheureusement due à du bruit ou, de façon plus heureuse, due à une généralisation reflètant une variabilité autour d'un prototype de l'image de départ.

a	b
d	x

$$x = (2m^p - b^p)^{1/p}$$

a	b
x	y

$$x = (2m^p - a^p)^{1/p}$$
$$y = (2m^p - b^p)^{1/p}$$

a	y
x	z

$$x = (2m^p - a^p)^{1/p}$$

FIGURE 3 – En haut, deux cas de reconstruction totale d'un bloc grâce à la connaissance de l'ordonnancement montré par les différences de grisé, et des moyenne et puissance analogiques, m et p. Ces cas peuvent apparaître différemment selon les ordonnancements (voir figure 2) et les positions possibles pour un même ordonnancement. En bas, la valeur de l'un des trois pixels inconnus peut être calculée grâce à la même connaissance.

4.1 Reconstruction d'un bloc

Pour introduire le processus de reconstruction, plaçons-nous dans le cas où l'on veut reconstruire un bloc de 2×2 pixels dont on connaît un certain nombre de pixels. Selon les circonstances, la connaissance de l'ordonnancement, de la puissance et de la moyenne analogiques permet d'accroître ou pas le nombre de pixels connus dans le bloc par la détermination éventuelle de la valeur de certains pixels inconnus.[3]

Si trois pixels sur quatre sont connus, la connaissance de l'ordonnancement, de la puissance et de la moyenne analogiques permet de savoir si le pixel inconnu est un extrême ou un moyen. Cette connaissance fonde le calcul suivant qui donne la valeur du pixel inconnu.

$$m_p(v, x) = m \;\Rightarrow\; x = (2m^p - v^p)^{1/p} \qquad (3)$$

Ci-dessus, v désigne la valeur de l'autre extrême ou moyen ; m est la moyenne analogique et p la puissance. De trois pixels connus, on passe à quatre et le bloc est alors entièrement reconstruit. Ce cas est illustré par le premier bloc de la figure 3.

Si deux pixels sur quatre sont connus, et que la connaissance de l'ordonnancement indique que l'un est l'un des extrêmes et l'autre l'un des moyens, cet ordonnancement permet d'établir lequel des pixels inconnus est l'autre extrême, le second étant l'autre moyen. L'utilisation de la formule (3) permet alors le calcul des valeurs, comme illustré par le deuxième bloc de la figure 3.

Maintenant, si les deux pixels connus sont les deux extrêmes, on ne peut accroître le nombre de pixels connus et l'on ne peut reconstruire exactement le bloc car il existe une infinité de possibilités pour les moyens. Il en est de même en échangeant extrêmes par moyens dans la phrase précédente.

3. Par simplification, l'exposé qui suit passe sous silence les cas particuliers où, par exemple, la puissance analogique est égale à $\pm\infty$.

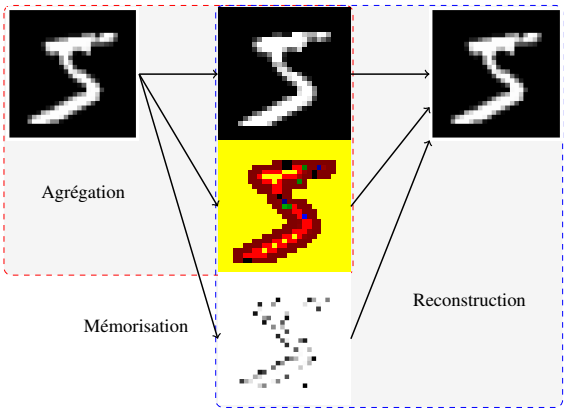

FIGURE 4 – L'image de gauche est soumise à l'agrégation par analogie. Il en résulte deux images réduites des moyennes analogiques, et des puissances analogiques selon les ordonnancements. Ces informations sont utilisées dans la reconstruction à droite, en combinaison avec un ensemble de pixels mémorisés à partir de l'image de gauche.

Si un seul pixel sur quatre est connu, comme illustré par le troisième bloc de la figure 3, la connaissance de l'ordonnancement permet de déterminer où trouver l'autre pixel, de valeur forcément inconnue, de la même classe, extrêmes ou moyens, que le pixel connu. L'équation (3) permet alors de calculer la valeur de ce pixel inconnu. On voit donc que le nombre de pixels connus dans le bloc s'accroît d'un pixel.

Il résulte de ce qui vient d'être dit que la connaissance produite par l'agrégation par analogie permet d'accroître le nombre de pixels connus dans certains blocs. Cette propriété est exploitée dans la reconstruction d'une image entière.

4.2 Reconstruction d'une image entière

Algorithme. Un algorithme de reconstruction d'image procédera en scannant l'image entière bloc par bloc. Chaque bloc à reconstruire se voit affecter un score correspondant au nombre de pixels qui y resteront inconnus après la reconstruction des pixels inconnus selon les règles données dans l'alinéa précédent.

La reconstruction complète se fera par itération. Au début tous les blocs sont présents dans la liste. À chaque étape d'itération, le score de chacun des blocs de la liste est calculé, la liste est triée par ordre croissant de score, les pixels reconstructibles sont reconstruits selon l'ordre des blocs dans la liste et les blocs entièrement reconstruits sont éliminés de la liste. On voit que la présence et l'ordre des blocs dans la liste change à cause des pixels reconstruits partagés par les blocs. La liste devient vide au bout d'un certain temps, ce qui signifie que l'image entière a été reconstruite. L'algorithme s'arrête alors.[4]

4. Nous renvoyons à l'article [10] pour une version précédente de l'algorithme, qui avait le désavantage de ne pas assurer que la liste se vide, d'où reconstruction avec ambiguïté. Ici la liste devient vide car on atteint toujours une configuration où tous les pixels sur les bords sont connus. Il est aisé de montrer (figure 4 de l'article cité) que la connaissance de deux bords perpendiculaires permet de reconstruire toute l'image.

Pixels mémorisées de l'image de départ. L'algorithme précédent requiert de connaître quelques pixels dans l'image à reconstruire pour pouvoir démarrer. On a vu plus haut qu'à chaque profondeur l'agrégation produisait deux images réduites, celle des moyennes analogiques, qui servira d'image de départ pour la profondeur suivante, et celles des puissances analogiques avec les ordonnancements. Mais ces informations ne suffisent pas à la reconstruction de l'image, car l'étendue des valeurs prises par les pixels d'une image ne peut se déduire de la donnée de ces seules données. On retient donc un certain nombre de pixels de l'image précédente pour aider à la reconstruction. On peut estimer le nombre de pixels suffisant à cette reconstruction à $2 \times n - 1$ pour une image carrée.[5] Nous tirons donc $2 \times n - 1$ pixels au hasard tels que leurs valeurs s'échelonnent uniformément entre les valeurs minimale et maximale des pixels de l'image de départ. La position de ces pixels dépend de l'image de départ et ils sont en pratique dispersés sur toute l'image. Pour résumer, la figure 4 illustre le processus de reconstruction d'une image à une profondeur donnée.

4.3 Reconstruction itérative

À chaque profondeur, l'image reconstruite à la profondeur précédente remplace l'image des moyennes produite à la profondeur courante lors de l'agrégation. Cette image reconstruite, combinée aux puissances et aux ordonnancements, ainsi qu'aux pixels mémorisés, permet de reconstruire une image, de taille plus grande, qui viendra à son tour remplacer l'image des moyennes à la profondeur suivante, en remontant vers l'image de départ.

5 Classification d'image

L'agrégation par analogie produit une image des moyennes analogiques qui atténue progressivement les détails sans, a priori, modifier la sémantique des classes. On peut s'interroger pour savoir si l'image reconstruite est toujours reconnaissable. Pour toute profondeur d'agrégation possible, nous nous proposons d'inspecter la qualité de reconstruction. Si les images appartiennent à des classes, cette interrogation s'interprète comme un problème de classification.

Buts de l'étude. Pour les images reconstruites à partir d'une certaine profondeur d'agrégation, les statistiques des caractéristiques des distributions divergent progressivement avec la profondeur d'agrégation-reconstruction des caractéristiques de distribution d'images de référence utilisées éventuellement dans une phase d'apprentissage. On peut s'interroger sur le fait d'avoir, sous l'effet de l'agrégation-reconstruction, un jeu de test hors-distribution correspondant à une situation de décalage de covariable.
On peut aussi se demander si la dérive des caractéristiques dans un jeu de test est cohérente avec la dérive observée dans un jeu d'entraînement. Et l'on peut donc envisager un protocole d'expériences dans lequel on effectuera des expériences de classification de tous les jeux de test produits par agrégation-reconstruction pour toutes les profondeurs possibles, à l'aide de tous les jeux d'entraînement produits par agrégation-reconstruction à toutes les profondeurs possibles.

Grille d'expériences. Après entraînement, chaque modèle peut être évalué non seulement sur l'ensemble de test correspondant de même profondeur mais aussi sur tous les autres ensembles de test à toutes les profondeurs possibles. En plus des données agrégées-reconstruites on utilisera bien sûr les données originales sans modification ; cela tient lieu de profondeur 0. Au total donc, si on appelle P la profondeur maximale, on construira $(1 + P)$ modèles correspondant au jeu d'entraînement original plus toutes les profondeurs sur ce même jeu. On mènera alors $(1 + P) \times (1 + P)$ expériences de classification pour chacune des $(1 + P)$ profondeurs d'agrégation-reconstruction du jeu de test.

Pour préciser, une expérience individuelle consiste à construire un modèle pour un classificateur donné sur un jeu d'entraînement obtenu pour une profondeur d'agrégation-reconstruction donnée. Le résultat de l'expérience individuelle est la qualité de classification du modèle sur un jeu de test obtenu pour une profondeur d'agrégation-reconstruction donnée, éventuellement différente de la profondeur du jeu d'entraînement du modèle. Nous utilisons simplement le pourcentage d'images classées correctement sur l'ensemble du jeu de test.

Analyse des résultats. La réalisation de toutes les expériences individuelles permettra de tracer une grille à deux axes : horizontalement selon la profondeur d'agrégation-reconstruction dans le jeu d'entraînement et verticalement selon la profondeur dans le jeu de test ; chaque case sera pour une expérience individuelle.

Une première vue sur cette grille consistera à examiner la qualité de classification à profondeur égale dans les jeux d'entraînement et de test afin d'étudier la pertinence de l'hypothèse de scénario de décalage de covariable. Idéalement, on voudrait que la qualité de classification soit constante et maximale sur cette diagonale.

On pourra aussi prendre sur la grille des coupes horizontales ou verticales et calculer la moyenne et l'écart-type des scores afin d'étudier, pour un jeu à profondeur fixée, la variation au travers de toutes les profondeurs dans l'autre jeu. Une coupe verticale, pour une profondeur fixée du jeu d'entraînement, permettra d'analyser à quel degré un jeu d'entraînement à profondeur fixé est capable de reconnaître des images agrégées-reconstruites à n'importe quelle profondeur. Pour chaque type de classificateur, c'est un test de sa sensibilité aux images agrégées-reconstruites. Sur une coupe horizontale, pour une profondeur fixée du jeu de test, l'écart-type des scores rendra compte de l'impact de l'agrégation-reconstruction au niveau sémantique dans le modèle, c'est-à-dire en quoi les images agrégées-reconstruites du jeu d'entraînement restent fidèles à leur classe.

5. C'est le nombre de pixels sur les bords haut et gauche, ce qui permet une reconstruction diagonale après diagonale depuis le coin en haut à gauche. En effet, une diagonale de pixels connu offre une configuration systématique de blocs de trois pixels connus pour un seul inconnu. On peut donc atteindre le coin en bas à droite (voir [10]).

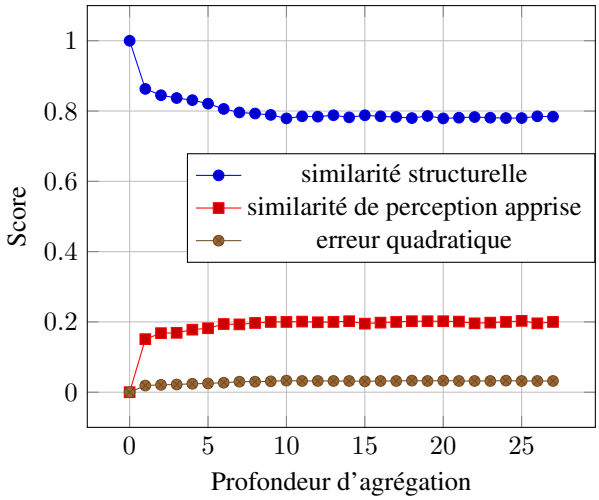

FIGURE 5 – Scores de qualité de reconstruction. Les scores à la profondeur 0 en abscisse donnent les scores de référence puisqu'ils résultent de la comparaison des images de départ avec elles-mêmes.

Types de classificateur. Afin de mesurer la sensibilité de différents types de modèles de classification à l'agrégation-reconstruction, on pourra répéter toutes les expériences précédentes en utilisant plusieurs méthodes de classification couramment utilisées et représentatives des approches traditionnelles.

6 Expériences

Données. Nos expériences utilisent MNIST [6], jeu de données bien connu en traitement d'image. Il contient 70 000 images de chiffres écrits à la main centrés et cadrés dans un carré de 28×28 pixels. Les pixels ont des valeurs de gris du blanc au noir entre 0 et 255. Étant donnée la taille des images, la profondeur maximale d'agrégation par analogie est de 27, ce qui correspond à une image finale réduite à un seul pixel.

6.1 Reconstruction d'image

Conditions expérimentales. Nous reconstruisons toutes les images du jeu de données à partir des informations produites par agrégation par analogie pour toutes les profondeurs possibles de 1 à 27. Notre objectif est de comparer les images reconstruites aux images de départ.

Évaluation. Les indices suivants sont utilisés pour évaluer la qualité des images reconstruites.

— L'indice de similarité structurelle [15] évalue la similarité visuelle entre deux images. Il est supposé refléter la perception humaine. Des valeurs élevées indiquent une similarité accrue.

— La similarité de perception apprise [18] reflète la différence de perception entre deux images, à partir de la similarité des représentations des blocs les constituant dans un espace latent issu d'un apprentissage. Comme son nom ne l'indique pas, des valeurs faibles sont préférables.

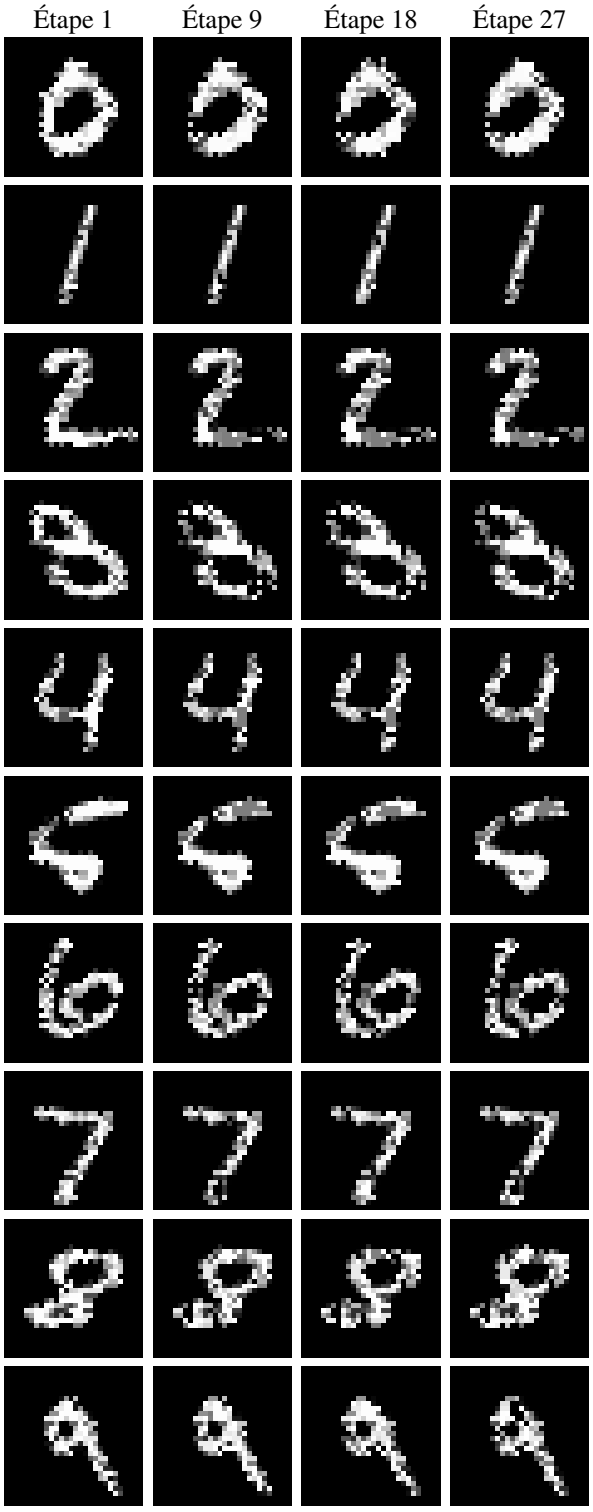

FIGURE 6 – Reconstruction d'images MNIST à partir de différentes étapes d'agrégation par analogie. Exemple tiré au hasard pour chaque chiffre de 0 à 9.

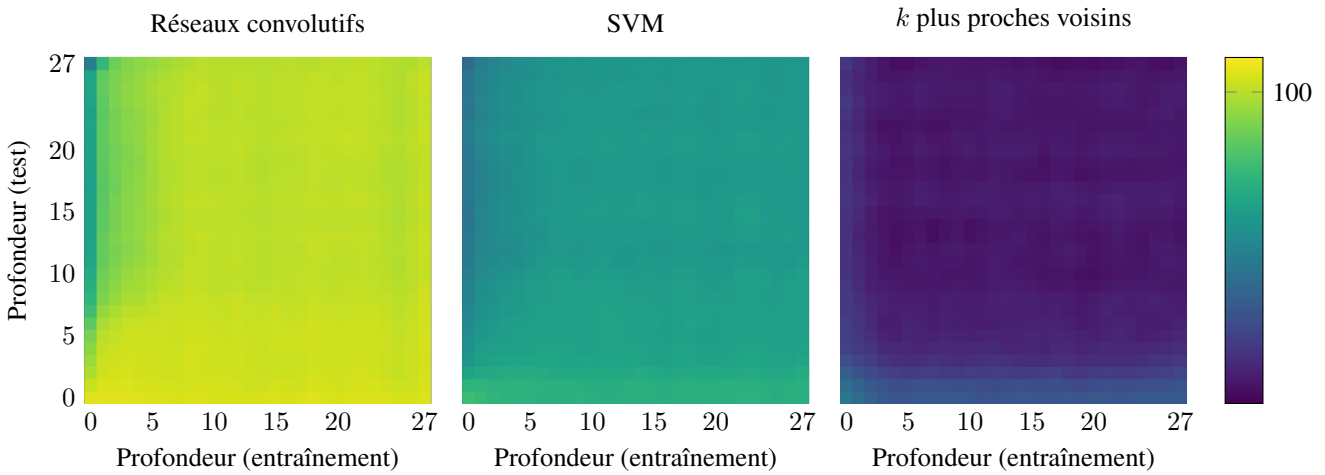

FIGURE 7 – Scores de classification pour différentes profondeurs d'agrégation sur les jeux d'entraînement (axe horizontal) et de test (axe vertical). La profondeur d'agrégation-reconstruction varie de 0 (images originales) à 27 (images réduites à 1 pixel puis reconstruites).

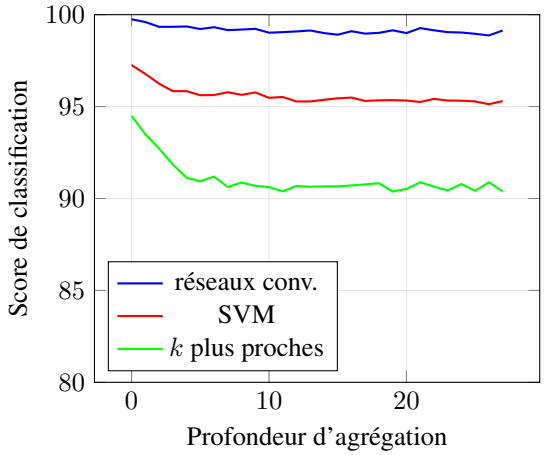

FIGURE 8 – Exactitude de la classification à profondeur identique dans les jeux d'entraînement et de test

— L'erreur quadratique en moyenne est le plus simple des indices. C'est la moyenne des différences au carré pixel par pixel entre deux images, ici l'image reconstruite et l'image de départ. Puisque c'est une erreur, de faibles valeurs indiquent une meilleure qualité de reconstruction.

Nous normalisons les trois indices précédents entre 0 et 1 afin de pouvoir les présenter dans un même graphe.

Résultats. La figure 5 présentent les valeurs des indices précédents en moyenne sur toutes les images pour toutes les profondeurs. On observe, pour la similarité de perception apprise et l'erreur quadratique en moyenne, une augmentation dès l'étape 1, ce qui reflète une certaine différence d'avec l'image de départ. Mais on observe ensuite une étonnante stabilité jusqu'à la profondeur maximale. L'indice de similarité structurelle détecte plus finement la pente aux premières profondeurs, mais il met en évidence la stabilité et la convergence exhibée par les deux autres indices à partir de la profondeur 10.

Des images reconstruites sont présentées dans la figure 6 pour quatre profondeurs d'agrégation-reconstruction uniformément réparties. Il faut noter que, pour l'étape 27, on part d'un seul pixel et l'on remonte jusqu'à l'image de départ à l'aide des informations mémorisées à chaque profondeur intermédiaire lors de l'agrégation.

Interprétation. Du fait de cette grande qualité de reconstruction, on est enclin à conclure que l'agrégation par analogie constitue bien une opération d'abstraction cohérente. Il est cependant nécessaire de répéter que l'agrégation par analogie n'est pas une compression. Le volume d'information produit est plus grand que dans la seule image de départ. Mais, même si elle provient entièrement de l'image de départ, et mise à part l'image réduite des moyennes analogiques, l'information est d'une nature bien différente. L'éclatement de l'information explique sans doute la qualité des images obtenues, qualité perceptible à l'œil nu et confirmée par les indices automatiques. On peut donc se demander si une quelconque généralisation a été obtenue et dans quelle mesure les variations entre les images reconstruites et les images de départ sont cohérentes avec ce que représentent les images, c'est-à-dire leur classe, ici le chiffre représenté par l'image. Cette question est précisément celle de la classification de ces images abordée ci-dessous.

6.2 Classification d'image

Afin de poursuivre l'examen des différences entre images reconstruites et images de départ d'un point de vue plus sémantique, nous précisons ci-dessous les expériences de classification dont la méthodologie générale a été décrite plus haut.

Classificateurs. Trois classificateurs sont choisis pour leur caractère classique et leur représentativité. Ils diffèrent grandement en termes de performance absolue. Nous les utilisons justement comme sondes pour analyser les effets structurels de l'agrégation-reconstruction sur le comporte-

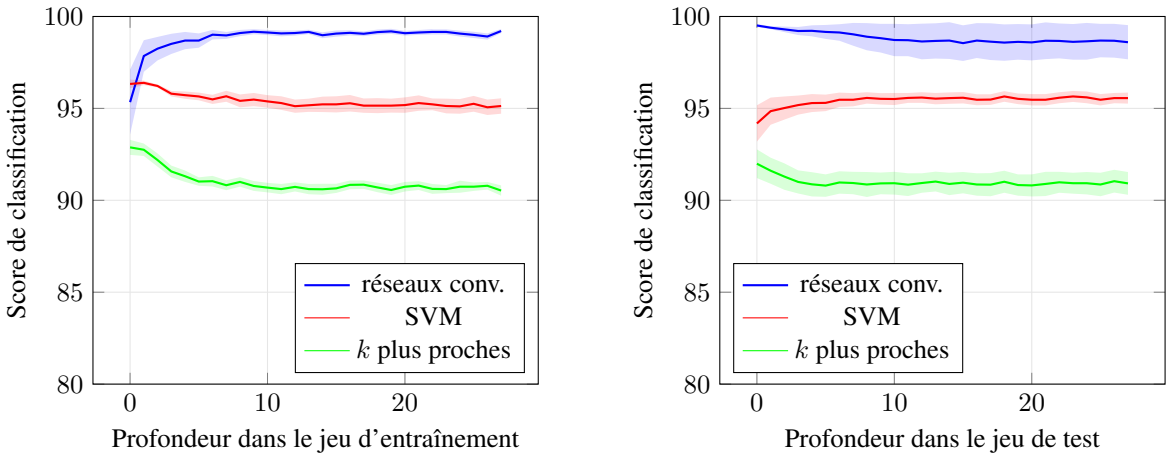

FIGURE 9 – À gauche, moyennes et écart-types de chaque coupe verticale dans les graphes de la figure 7, c'est-à-dire qualité de classification en moyenne et écart-type sur tous les jeux de test de toutes les profondeurs possibles pour les trois méthodes pour chaque modèle individuel créé à une profondeur fixée du jeu d'entraînement. À droite, moyennes et écart-types de chaque coupe horizontale dans les graphes de la figure 7, c'est-à-dire qualité de classification de tous les modèles, pour chaque profondeur individuelle d'agrégation-reconstruction du jeu de test.

ment de l'apprentissage en aval. Ce sont les suivants. Pour chacun d'eux, les images sont aplaties en vecteurs unidimensionnels avant usage.

- Réseaux de neurones convolutifs : nous adoptons le même réseau que dans [1] et l'entraînons sur 150 époques. Le modèle se compose de cinq couches convolutives qui réduisent progressivement la taille de la carte de caractéristiques tout en augmentant le nombre de canaux, suivies d'une couche entièrement connectée pour la classification.
- Système à vastes marges (SVM) [2] : nous utilisons un noyau de fonctions de base radiales.
- Méthode des k plus proches voisins [14] : nous posons $k = 3$.

Résultats. La figure 7 donne les grilles de qualité de classification pour les trois types de classificateur pour toutes les profondeurs dans les jeux d'entraînement et de test. Rappelons qu'il s'agit du pourcentage d'images classées correctement sur l'ensemble du jeu de test. Bien que pas forcément visibles, les scores maximaux par colonne sont tous situés en bas des grilles pour chacun des classificateurs utilisés. De façon semblable, les scores minimaux par ligne sont tous situés à gauche de la grille, sauf pour les k plus proches voisins où se sont les scores maximaux par ligne qui se trouvent à gauche.

La figure 8 présente les scores pour des profondeurs identiques dans les jeux d'entraînement et de test. Il s'agit d'expériences de classification dans lesquelles les images classifiées sont des images reconstruites à partir de la même profondeur que les images utilisées dans l'entraînement. On observe d'abord une dégradation de l'exactitude de classification jusqu'à environ la profondeur 5 et une stabilisation ensuite. Dans le cas des réseaux de neurones, les scores restent très comparables (1 % de différence), la dégradation est légère pour les SVM (3 % environ), mais plus prononcée avec la méthode des k plus proches voisins (5 %). Ces graphes confirment que l'agrégation suivie de reconstruction introduit une certaine variabilité dans les données, ais cette variabilité est sans doute introduite dans les premières profondeurs et pas vraiment après. Cette supposition sera confirmée par les analyses qui suivent.

La graphe de gauche de la figure 9 donne les mesures en coupes verticales dans les grilles. Il s'agit pour chaque modèle entraîné sur les données d'une profondeur particulière, de sa performance en moyenne sur toutes les images de test à toutes les profondeurs. Ces courbes montrent donc la capacité d'images à une profondeur donnée de permettre de classifier des images de n'importe quelle profondeur. Pour les classificateurs SVM et k plus proche voisins, après une légère décroissance, les courbes stagnent. Le comportement des réseaux de neurones est différent : le modèle entraîné sur les images de départ n'arrive pas classifier très bien des images de n'importe quelle profondeur, mais toujours mieux que la méthode des plus proches voisins et de façon comparable à la méthode SVM. Mais la pente remonte très vite et les scores de classification se stabilisent immédiatement à leurs plus hautes valeurs.

Le graphe de droite de la figure 9 donne les mesures en coupes horizontales dans les grilles. Pour un jeu de test d'une profondeur donnée, on lit la moyennes des scores de classification obtenus avec tous les modèles à toutes les profondeurs. Ces courbes montrent donc la capacité qu'ont des images d'une certaine profondeur à être reconnues pour ce qu'elles sont par des classificateurs entraînés à des profondeurs diverses. On observe, selon les méthodes, une décroissance ou une croissance jusqu'à la profondeur 5, suivie dans tous les cas d'une stagnation. L'ampleur de la croissance ou décroissance n'est pas énorme, mais elle reflète sans doute une dérive dans la distribution des images du jeu de test aux premières profondeurs. On peut constater que

l'écart-type pour les réseaux de neurones augmente avec le profondeur dans le jeu de test.

Interprétation. Les divers graphes présentés plus haut montrent des classements en score similaires pour les trois types de classificateurs utilisés. On retrouve le classement intuitif en performance : réseaux de neurones suivis de SVM, suivis des k plus proches voisins. Cette cohérence indique que tous les phénomènes observés ci-dessus résultent principalement du mécanisme d'agrégation-reconstruction étudié et pas d'architectures spécifiques.

On peut considérer que les résultats à profondeur égale sont en accord avec ceux obtenus en reconstruction seule. Ils démontrent une forte cohérence de l'agrégation-reconstruction mais la méthode des k plus proches voisins se révèle plus sensible à la dispersion des images aux premières profondeurs. Certaines des images se mettent sans doute immédiatement à ressembler à des images de classes voisines.

Un faisceau d'indices convergents permet d'aborder la question de la généralisation. Ce sont le comportement des scores sur les bords gauches et bas dans les grilles de la figure 7, et la variation jusqu'à la profondeur 5 dans les graphes de gauche et de droite de la figure 9, puis la stagnation observée après. Ce faisceau d'indices semble indiquer que les modèles qui utilisent des données d'entraînement d'une certaine profondeur à partir d'un certain seuil, sont capables de classer relativement bien les données de jeux de test de plus faible profondeur. Pour résumer, des images agrégées-reconstruites par analogie à une certaine profondeur, mais à partir d'un certain seuil, produisent des modèles capables de généraliser relativement bien à des images de n'importe quelle profondeur d'agrégation-reconstruction.

7 Conclusion

Dans cet article, nous avons proposé une méthode d'agrégation fondée sur l'analogie numérique, l'agrégation par analogie, et l'avons utilisée pour réduire des images, les reconstruire et les classifier. L'exposé de l'agrégation par analogie a montré comment l'opération proposée éclate l'information en, d'une part, une information semblable aux types de données initiales et, d'autre part, des informations bien spécifiques à l'analogie numérique. L'exposé du processus de reconstruction a montré comment ces données spécifiques sont recombinées pour reconstruire des images qui, d'après les résultats d'expériences, semblent de qualité relativement bonnes. Nous nous sommes alors interrogé sur la conservation de la sémantique des images reconstruites sous l'angle de la classification. Nos expériences laissent entrevoir le fait que les données produites par agrégation par analogie puis reconstruction permettent de créer des modèles capables d'un certain degré de généralisation au travers des profondeurs de reconstruction (après un certain seuil).

L'analogie numérique s'avère prometteuse comme méthode d'abstraction car elle semble non seulement cohérente (les scores de classification des images reconstruites à une certaine profondeur sont classifiées relativement correctement par des classificateurs entraînés sur des images reconstruites à partir de la même profondeur), mais encore capable de généralisation (les scores de classification des images reconstruites à une certaine profondeur sont relativement bien classifiées par tous les classificateurs entraînés sur des images reconstruites à n'importe quelle profondeur).

La conclusion à tirer des expériences présentées ici est que le processus d'agrégation par analogie suivie de la reconstruction apparaît bien comme une opération d'abstraction qui semble conduire à une certaine généralisation.

Une part de nos travaux futurs continuera dans la lignée des expériences présentées ici. Pour ce qui est de la reconstruction d'image, nous étudions déjà le nombre de pixels mémorisés nécessaires à une reconstruction exacte, voire leur nécessité même, en lien avec la nécessité de la mémorisation des puissances analogiques. Bien que les expériences présentées ici utilisassent des images à un seul canal (gris), nous menons aussi des expériences sur des images à plusieurs canaux (couleurs).

Nos travaux futurs s'intéresseront aussi à d'autres tâches en aval. Notre but à long terme est d'explorer l'application de l'analogie numérique au raisonnement par analogie et d'explorer comment les informations abstraites extraites par l'analogie numérique peuvent contribuer au processus d'abstraction. En lien avec le traitement d'image, et dans le but de résoudre des tâches complexes, comme les différentes déclinaisons de matrices progressives de Raven (RAVEN, ARC ou ConceptARC), nous abordons déjà le calcul symbolique d'analogies par la résolution directe d'analogies entre chiffres manuscrits, au niveau des images .

Remerciements

Les travaux présentés ici ont bénéficié d'une subvention de recherche de l'université Waseda, 2025R-034, intitulée « Analogie numérique : formalisation mathématique et application pratique à l'apprentissage automatique ».

Ces travaux ont aussi été partiellement soutenus par le projet ANR-22-CE23-0023 « *Analogies : from theory to tools and applications* (AT2TA) ».

Références

[1] Sanghyeon An, Minjun Lee, Sanglee Park, Heerin Yang, and Jungmin So. An ensemble of simple convolutional neural network models for MNIST digit recognition, 2020.

[2] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3) :273–297, 1995.

[3] Francisco Cunha, Yves Lepage, Zied Bouraoui, and Miguel Couceiro. Generalizing analogical inference from Boolean to continuous domains. In *Proceedings of the 40th Annual AAAI Conference on Artificial Intelligence (AAAI 2026)*, Singapore, jan 2026. AAAI Press. To appear.

[4] Miguel de Carvalho. Mean, what do you mean ? *The American Statistician*, 70(3) :270–274, 2016.

[5] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11) :2278–2324, 1998.

[6] Yann LeCun, Bernhard Boser, John S. Denker, Donnie Henderson, Richard E. Howard, Wayne Hubbard, and Lawrence D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4) :541–551., 1989.

[7] Chen-Yu Lee, Patrick W. Gallagher, and Zhuowen Tu. Generalizing pooling functions in convolutional neural networks : Mixed, gated, and tree. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics (AISTATS), JMR : W&CP*, volume 51, pages 464–472, Cadiz, Spain, 2016.

[8] Yves Lepage and Miguel Couceiro. Analogie et moyenne généralisée. In Jean-Guy Mailly, François Schwarzentruber, and Anaelle Wilczynski, editors, *Actes des 18es Journées d'Intelligence Artificielle Fondamentale et des 19es Journées Francophones sur la Planification, la Décision et l'Apprentissage pour la conduite de systèmes (JIAF 2024 et JFPDA 2024)*, pages 114–124, La Rochelle, France, Juillet 2024.

[9] Yves Lepage and Miguel Couceiro. L'analogie numérique revisitée et étendue, précisée, rétrécie ou agrandie. In Jean-Guy Mailly, François Schwarzentruber, and Anaelle Wilczynski, editors, *Actes des 19es Journées d'Intelligence Artificielle Fondamentale et des 20es Journées Francophones sur la Planification, la Décision et l'Apprentissage pour la conduite de systèmes (JIAF 2025 et JFPDA 2025)*, pages 1–12, Dijon, France, Juillet 2025.

[10] Jakub Pilimon, Miguel Couceiro, and Yves Lepage. Analogical pooling for image reconstruction. In Miguel Couceiro Zied Bouraoui, editor, *Proceedings of the 4th Workshop on the Interactions between Analogical Reasoning and Machine Learning (IARML 2025), co-located with IJCAI 2025*, pages 27–41. HAL, Montréal (Québec), Canada, 2025.

[11] J.C. Raven. *The Performances of Related Individuals in Tests Mainly Educative and Mainly Reproductive Mental Tests Used in Genetic Studies*. University of London (King's College), 1936.

[12] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net : Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015)*, pages 234–241, Cham, 2015. Springer International Publishing.

[13] Assaf Shmuel, Oren Glickman, and Teddy Lazebnik. Machine and deep learning performance in out-of-distribution regressions. *Machine Learning : Science and Technology*, 5, 01 2025.

[14] Peter E. Hart Thomas M. Cover. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1) :21–27, 1967.

[15] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment : from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4) :600–612, 2004.

[16] Huaxiu Yao, Caroline Choi, Bochuan Cao, Yoonho Lee, Pang Wei Koh, and Chelsea Finn. Wild-time : a benchmark of in-the-wild distribution shift over time. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA, 2022. Curran Associates Inc.

[17] Jinliang Yao, Yongqing Li, Bing Yang, and Chenrui Wang. Learning global image representation with generalized-mean pooling and smoothed average precision for large-scale CBIR. *IET Image Processing*, 17(9) :2748–2763, 2023.

[18] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2018)*, pages 586–595, Salt Lake City, Utah, 2018.