

APPRENTISSAGE CONTINU POUR LA CLASSIFICATION DES CONSOMMABLES DE SANTÉ

Mohamed Demes et Taher Lmouden

ABSTRACT

Ces dernières années, répondre aux défis économiques liés au recyclage des dispositifs médicaux jetables est devenu de plus en plus crucial, si bien que le processus de recyclage a évolué. Pourtant, les solutions actuelles présentent de nombreux défis, le plus notable étant la forte implication de la main-d'œuvre, menant à des pertes de temps et d'argent, ce qui rend ces solutions inefficaces.

Parmi les solutions innovantes qu'on peut adopter, l'intégration de l'intelligence artificielle et l'Internet des objets (IoT) pour automatiser le processus du recyclage. Notre objectif est de créer un système qui utilise des caméras pour détecter les déchets médicaux recyclables. Cependant, des nouveaux types de déchets apparaissent continuellement, ce qui fait qu'on doit garantir que notre système évolue constamment, justifiant le recours à l'approche d'apprentissage continu pour leur détection, en s'appuyant sur le 'few shot learning', une méthode qui tire parti d'un petit nombre d'exemples pour entraîner notre modèle à généraliser vers de nouvelles données. D'où l'utilisation des modèles de fondation comme support; des modèles qui servent de base pour l'adaptation vers des tâches spécifiques avec un nombre limité d'exemples, offrant ainsi une méthode efficace pour l'apprentissage avec peu de données.

Mots clés: Foundation models, Internet des objets, Few shots.

1 INTRODUCTION

En examinant l'évolution constante des quantités de déchets produits par le secteur de la santé, nous pouvons évaluer les impacts économiques des consommables médicaux et l'importance de leur gestion à l'avenir. Selon le Conseil de recyclage des plastiques de santé, les coûts associés aux déchets de soins de santé étaient estimés à 36 milliards de dollars en 2020 à l'échelle mondiale et devraient atteindre 55 milliards de dollars en 2025. La solution actuellement établie pour limiter ces dépenses considérables est le recyclage des consommables de santé. Toutefois, le processus de recyclage comporte plusieurs défis, principalement liés à la dépendance envers l'intervention humaine. Cette dépendance implique des risques d'erreurs émises par l'humain, et des risques de blessures puisqu'ils sont souvent en contact avec des déchets potentiellement dangereux, les exposant ainsi à des effets toxiques et à des maladies contagieuses.

La solution que nous proposons repose sur l'utilisation de l'intelligence artificielle et de l'Internet des Objets (IoT) pour automatiser le processus de recyclage. Cette approche réduit la dépendance humaine en privilégiant la technologie et l'intelligence artificielle, diminuant ainsi les risques pour les ouvriers. Pour atteindre cet objectif, nous envisageons de développer un modèle capable de détecter et de classifier les déchets, afin de prendre des décisions de recyclage adaptées à chaque catégorie. Toutefois, la diversité croissante des déchets médicaux nécessite une adaptation continue du système. Pour répondre à cette exigence et concevoir un modèle autonome et durable, nous envisageons d'adopter une approche d'apprentissage continu. Cette approche vise à créer des modèles capables d'apprendre et de reconnaître de nouveaux types de déchets qui pourraient apparaître à l'avenir, en utilisant un minimum d'échantillons grâce à l'apprentissage "few-shot". Ils luttent également contre l'oubli catastrophique (Catastrophic forgetting) tout en maintenant leur efficacité dans la détection des déchets déjà connus. Cette adaptation continue garantit que nos systèmes demeurent pertinents et performants. Dans notre solution, nous exploiterons les capacités des modèles de fondation en matière d'apprentissage en few-shot et de leur adaptation pour contrer l'oubli catastrophique, afin de réaliser un apprentissage continu.

2 ÉTUDE DE L'EXISTANT

2.1 MODÈLES DE FONDATION

Les modèles de fondation (Foundation Models) sont une classe de modèles d'apprentissage automatique qui sont pré-entraînés sur une vaste quantité de données générales avant d'être affinés sur des tâches spécifiques. Ces modèles sont reconnus par leur capacité à générer des performances impressionnantes sur une large gamme de tâches d'apprentissage profond, y compris la compréhension du langage naturel, la vision par ordinateur, et plus encore. Les modèles de fondation reposent sur trois principes fondamentaux. Premièrement, la flexibilité et la généralisation qui signifie que les modèles de fondations sont conçus pour fonctionner sur diverses tâches sans nécessiter de modifications architecturales majeures. Par exemple, le même modèle peut traiter à la fois le texte, l'image, les tâches multimédia, etc.

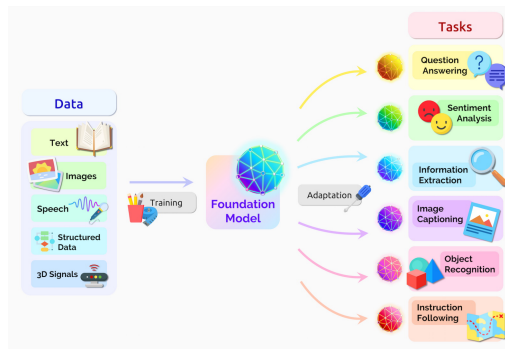


Figure 1: L'adaptation des modèles de fondation à différentes tâches.

Deuxièmement, l'échelle (ou "scaling") implique l'utilisation de grandes quantités de données pour l'entraînement du modèle et de grandes ressources technologiques. Cela signifie que les modèles de fondation sont conçus pour fonctionner efficacement sur des ensembles de données massives et à s'adapter à des infrastructures informatiques de grande envergure pour traiter et analyser ces données volumineuses(1).

Troisièmement, le transfert d'apprentissage (ou "transfer learning" en anglais) est une technique d'apprentissage automatique permettant d'initialiser le modèle avec des connaissances préalablement acquises au lieu de démarrer l'apprentissage à partir de zéro. En d'autres termes, le modèle est pré-entraîné sur une tâche source qui partage des caractéristiques ou des données similaires avec la tâche cible. Ensuite, les connaissances acquises par le modèle, sous forme de poids de réseau neuronal et de représentations apprises, sont transférées à un nouveau modèle destiné à la tâche cible, ce qui peut accélérer le processus d'apprentissage, améliorer les performances du modèle final, et adapter le modèle à des tâches spécifiques avec relativement peu de données supplémentaires pour le fine-tuning(1).

Grâce au transfert d'apprentissage, les modèles de fondation ont démontré une capacité d'apprentissage en few-shot tout en conservant une précision élevée(5). L'apprentissage en few-shot, qui signifie la capacité d'un modèle à apprendre à partir d'un minimum d'échantillons, est une technique qui a gagné en popularité ces dernières années, notamment grâce aux avancées dans le domaine de l'apprentissage automatique et de l'intelligence artificielle. Cette approche sera particulièrement pertinente pour notre tâche cible, car les nouvelles données à apprendre pour notre modèle auront des quantités de données limitées. Par conséquent, il est crucial de garantir une précision maximale avec un minimum d'échantillons.

2.2 APPRENTISSAGE CONTINU

L'apprentissage continu, également appelé "continual learning", vise à développer des algorithmes capables d'acquérir de nouvelles connaissances de manière continue tout en préservant celles déjà acquises. C'est crucial dans les systèmes d'IA qui doivent s'adapter à un flux constant de nouvelles données ou de nouvelles tâches sans oublier les informations précédemment apprises. La formation incrémentielle est au cœur de cette approche, permettant aux modèles d'évoluer progressivement avec de nouvelles données. Cependant, la formation progressive des modèles pose un défi car ils ont tendance à ajuster excessivement les paramètres sur les données actuelles, ce qui peut entraîner l'oubli des informations précédemment apprises. Ce phénomène est appelé "oubli catastrophique" et demeure un problème de recherche ouvert(6). Les techniques d'apprentissage continu sont utiles dans deux cas principaux :

- Adaptation rapide aux nouvelles données : Parfois, les modèles de machine learning doivent être régulièrement actualisés pour rester efficaces. Par exemple, un modèle détectant la fraude dans les transactions bancaires doit constamment apprendre de nouvelles techniques frauduleuses pour détecter rapidement les activités malveillantes.
- Personnalisation du modèle : Dans certains cas, les besoins de chaque utilisateur peuvent varier. Par exemple, dans un système de classification de documents, chaque utilisateur peut travailler avec des documents utilisant des termes et des styles différents. L'apprentissage continu permet d'adapter automatiquement le modèle à chaque utilisateur en utilisant ses propres données, assurant ainsi des résultats précis et adaptés à ses besoins spécifiques.

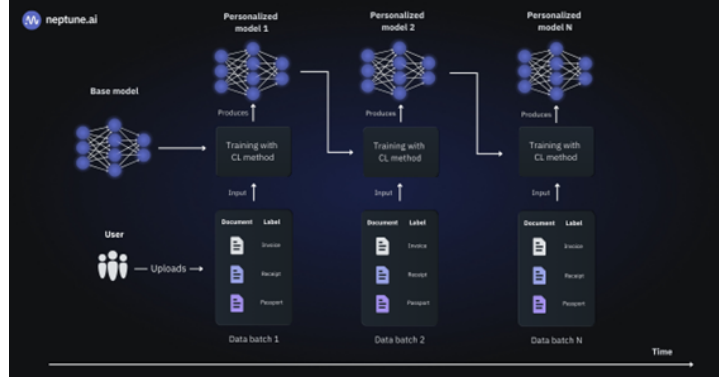


Figure 2: L'apprentissage continu pour la personnalisation du modèle.

On distingue trois principales approches dans l'apprentissage continu. Premièrement, les approches architecturales qui consistent à modifier la structure du modèle afin de lui permettre d'apprendre de nouvelles tâches sans oublier les précédentes. Par exemple, dans un scénario où un modèle doit être entraîné sur plusieurs tâches, différentes couches de classification peuvent être ajoutées au modèle de base pour chaque tâche spécifique. Cette approche permet de préserver les connaissances acquises tout en adaptant le modèle aux nouvelles données. Deuxièmement, la régularisation qui maintiennent la structure du modèle fixe pendant l'apprentissage, mais utilisent des techniques pour éviter l'oubli catastrophique. Par exemple, la distillation des connaissances consiste à transférer les connaissances d'un modèle pré-entraîné vers un modèle en cours d'apprentissage, afin de conserver les informations importantes, en incluant la modification des fonctions de perte et la sélection des paramètres à mettre à jour lors de l'apprentissage. Troisièmement, la méthode d'apprentissage continu basée sur la mémoire, implique de sauvegarder une partie des échantillons d'entrée dans une mémoire tampon pendant la formation, et à les réutiliser lors de l'apprentissage de nouvelles tâches. À titre d'illustration, le modèle sera entraîné sur un lot de données actuelles, ainsi que sur des échantillons sélectionnés de manière aléatoire dans la mémoire tampon. Cette approche permet de maintenir la diversité des données et d'éviter l'oubli catastrophique.

3 PARAMÈTRES

3.1 DATASET DESCRIPTION

Le jeu de données utilisé pour les différentes expériences mises en œuvre est intitulé "Medical-Waste-4.0-Dataset", créé par la Région de Toscane. Il a été élaboré dans le but de constituer une ressource précieuse pour la conception et le test de méthodes de vision par ordinateur dédiées au tri primaire des déchets médicaux. Ce jeu de données contient 4245 images de résolution 1920 x 1080 réparties en 13 classes. Un échantillon représentatif de chaque classe est illustré dans la figure 3.

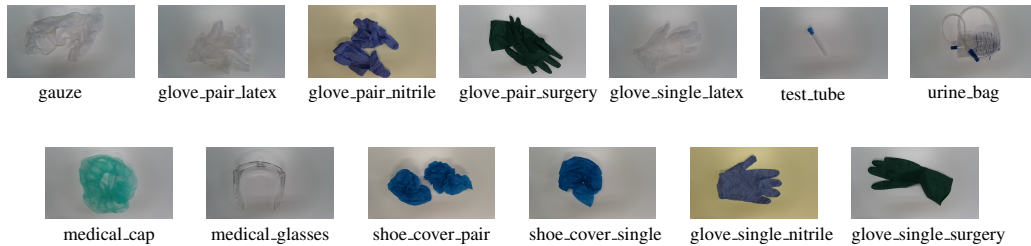


Figure 3: Echantillons de chaque classe du jeu de données.

Afin d'explorer les performances des modèles de base et les comparer à notre modèle de réseau de neurones simple, nous avons opté pour la division de notre jeu de données en trois ensembles distincts : l'ensemble d'entraînement, représentant 70% des échantillons (soit 2967 images), l'ensemble de validation, comprenant 15% des échantillons (soit 629 images), et enfin l'ensemble de test, comprenant 15% (soit 649 images).

3.2 MODELS DESCRIPTION

3.2.1 RÉSEAU DE NEURONES SIMPLE

Le réseau de neurones convolutif présenté est un modèle simple conçu pour la classification d'images. Il se compose de blocs de convolution suivis de couches de pooling pour extraire des caractéristiques, puis ces caractéristiques sont aplaties et transmises à des couches entièrement connectées pour la classification. La fonction d'activation ReLU est utilisée pour introduire de la non-linéarité dans le modèle. Ce modèle a été utilisé pour comparer la robustesse à l'apprentissage continu d'un modèle simple à celle des modèles de fondations.

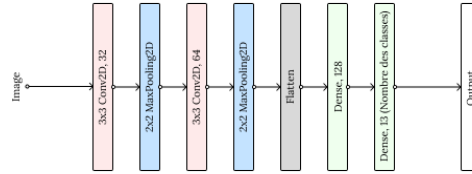


Figure 4: Architecture du modèle de réseau de neurones simple.

3.2.2 RESNET50

ResNet-50, un modèle de réseau neuronal convolutif (CNN) profond développé par Microsoft Research, fait partie de la famille des réseaux résiduels (ResNet), qui utilisent des connexions résiduelles pour faciliter l'entraînement de réseaux plus profonds. Avec ses 50 couches de convolution, ResNet-50 totalise 25,6 millions de paramètres. Ce modèle a été introduit dans le papier de recherche "Deep Residual Learning for Image Recognition"(4). Il est pré-entraîné sur le dataset ImageNet-1k, qui comprend 1 million d'images de résolution 224 x 224 réparties en 1000 classes. ResNet-50 est largement utilisé pour des tâches telles que la classification d'images et la détection d'objets, grâce à sa capacité à extraire des caractéristiques complexes à partir de données visuelles.

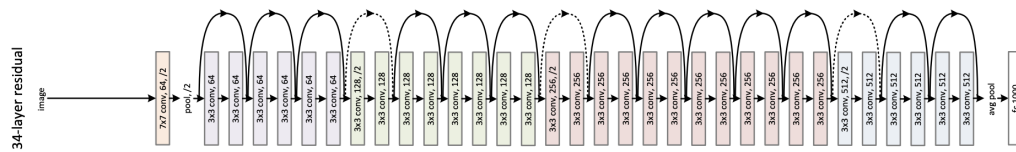


Figure 5: Architecture du modèle ResNet-50.

3.2.3 ViT

ViT, ou Vision Transformer, est un modèle révolutionnaire dans le domaine de la vision par ordinateur, développé par Google, qui applique les principes des transformeurs, à l'origine développés pour le traitement du langage naturel, à l'analyse d'images. Contrairement aux architectures CNN traditionnelles, ViT traite les images comme des séquences de jetons, ce qui lui permet de capturer des relations à longue portée entre les pixels. Avec ses 86,6 millions de paramètres, ViT a été introduit pour la première fois dans le papier de recherche intitulé "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale"(3). Il a été pré-entraîné sur une large collection d'images de manière supervisée, notamment ImageNet-21k(2), avec 14 millions d'images

de résolution de 224x224 pixels réparties en 21,843 classes. Ensuite, le modèle a été fine-tuné sur ImageNet 2012, un ensemble de données comprenant 1 million d'images et 1000 classes, également à une résolution de 224x224 pixels.

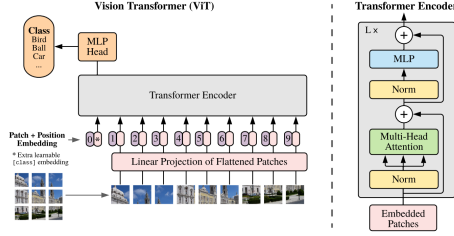


Figure 6: Architecture du modèle ViT.

4 EXPÉRIENCES

4.1 ÉVALUATION DU RÉSEAU DE NEURONES SIMPLE

Dans un premier temps, nous prévoyons d'évaluer un modèle de réseau de neurones simple. Pour ce faire, nous allons l'entraîner sur notre jeu de données d'entraînement pendant 10 époques, afin de l'utiliser comme référence de base pour la comparaison avec d'autres modèles fondamentaux. Ensuite, nous testerons ce modèle sur notre ensemble de test afin d'évaluer ses performances. Les résultats obtenus sont présentés dans les figures 7 et 8.

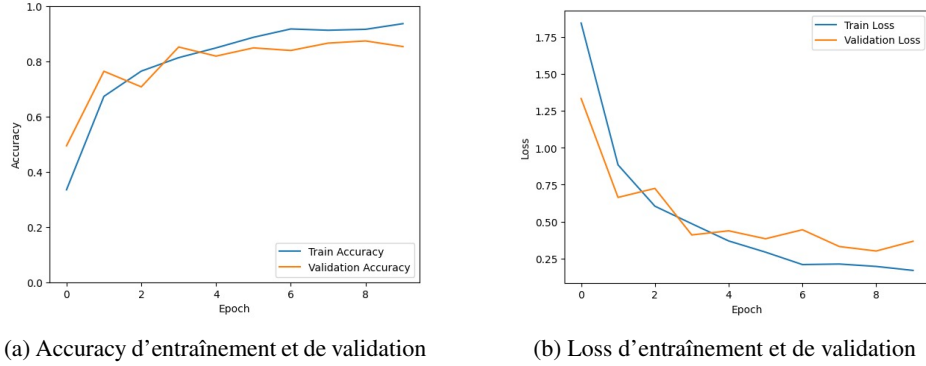


Figure 7: Accuracy et loss du modèle de réseau de neurones simple.

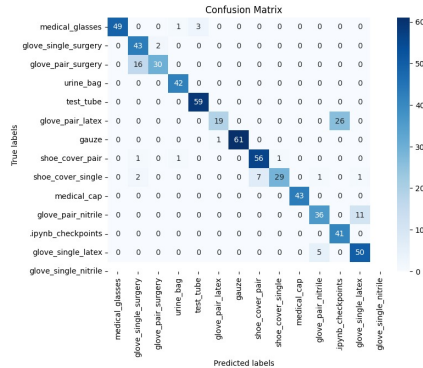


Figure 8: Matrice de confusion sur l'ensemble 'Testing' du modèle de réseau de neurones.

En examinant les graphiques d'accuracy et de perte sur les ensembles d'entraînement et de validation, nous avons constaté que le modèle atteignait de bonnes performances en dixième époque, avec une accuracy de 0.86 et une perte de 0.38 sur l'ensemble de test. Cette observation a été confirmée par la matrice de confusion, qui a témoigné de la capacité du modèle à prédire correctement les échantillons de test.

4.2 ÉVALUATION DES MODÈLES DE FONDATION SANS FINE-TUNING

Dans cette expérience, nous explorons la capacité de prédiction des modèles de fondation afin de déterminer s'ils peuvent correctement identifier nos classes. Pour obtenir des résultats plus tangibles, nous avons ajouté une catégorie nommée 'false prediction' à la matrice de confusion. Cette catégorie indique que le modèle a attribué des classes qui ne correspondent pas à celles du dataset utilisé.

4.2.1 RESNET-50

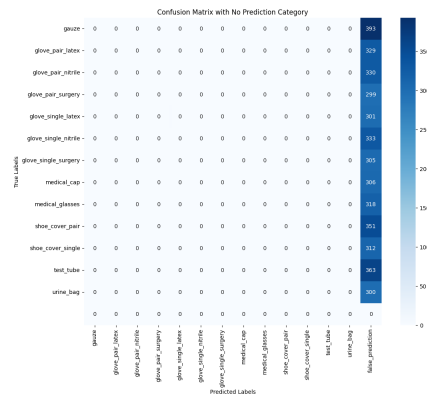


Figure 9: Matrice de confusion pour le jeu de données utilisant le modèle ResNet-50 non fine-tuné.

4.2.2 ViT

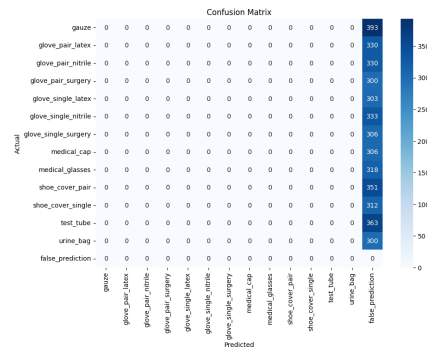


Figure 10: Matrice de confusion pour le jeu de données utilisant le modèle ViT non fine-tuné.

4.3 ÉVALUATION DES MODÈLES DE FONDATION APRÈS FINE-TUNING

- Décrire en quoi consistent les expériences, exemple : nous investiguons les performances d'un certain nombre de modèles de fondation pour la reconnaissance de déchets médicaux.

- Dans un premier temps, nous allons voir comment se comportent les modèles de fondation sur un dataset de déchets médicaux sans finetuning. L'idée est de voir ... (exemple: à quel point les modèles de fondation sont robustes à de nouvelles classes)
- Dans un deuxième temps, nous allons finetuner ces modèles sur notre dataset et voir comment ... (quel est le meilleur modèle, queles sont les qttés de données nécessaire)

4.4 RESULTS WITHOUT FINETUNING

4.5 RESULTS WITH FINETUNING

5 SUBMISSION OF CONFERENCE PAPERS TO ICLR 2024

ICLR requires electronic submissions, processed by <https://openreview.net/>. See ICLR’s website for more instructions.

If your paper is ultimately accepted, the statement `\iclrfinalcopy` should be inserted to adjust the format to the camera ready requirements.

The format for the submissions is a variant of the NeurIPS format. Please read carefully the instructions below, and follow them faithfully.

5.1 STYLE

Papers to be submitted to ICLR 2024 must be prepared according to the instructions presented here.

Authors are required to use the ICLR \LaTeX style files obtainable at the ICLR website. Please make sure you use the current files and not previous versions. Tweaking the style files may be grounds for rejection.

5.2 RETRIEVAL OF STYLE FILES

The style files for ICLR and other conference information are available online at:

<http://www.iclr.cc/>

The file `iclr2024_conference.pdf` contains these instructions and illustrates the various formatting requirements your ICLR paper must satisfy. Submissions must be made using \LaTeX and the style files `iclr2024_conference.sty` and `iclr2024_conference.bst` (to be used with \LaTeX 2e). The file `iclr2024_conference.tex` may be used as a “shell” for writing your paper. All you have to do is replace the author, title, abstract, and text of the paper with your own.

The formatting instructions contained in these style files are summarized in sections 6, 7, and 8 below.

6 GENERAL FORMATTING INSTRUCTIONS

The text must be confined within a rectangle 5.5 inches (33 picas) wide and 9 inches (54 picas) long. The left margin is 1.5 inch (9 picas). Use 10 point type with a vertical spacing of 11 points. Times New Roman is the preferred typeface throughout. Paragraphs are separated by 1/2 line space, with no indentation.

Paper title is 17 point, in small caps and left-aligned. All pages should start at 1 inch (6 picas) from the top of the page.

Authors’ names are set in boldface, and each name is placed above its corresponding address. The lead author’s name is to be listed first, and the co-authors’ names are set to follow. Authors sharing the same address can be on the same line.

Please pay special attention to the instructions in section 8 regarding figures, tables, acknowledgments, and references.

There will be a strict upper limit of 9 pages for the main text of the initial submission, with unlimited additional pages for citations.

7 HEADINGS: FIRST LEVEL

First level headings are in small caps, flush left and in point size 12. One line space before the first level heading and 1/2 line space after the first level heading.

7.1 HEADINGS: SECOND LEVEL

Second level headings are in small caps, flush left and in point size 10. One line space before the second level heading and 1/2 line space after the second level heading.

7.1.1 HEADINGS: THIRD LEVEL

Third level headings are in small caps, flush left and in point size 10. One line space before the third level heading and 1/2 line space after the third level heading.

8 CITATIONS, FIGURES, TABLES, REFERENCES

These instructions apply to everyone, regardless of the formatter being used.

8.1 CITATIONS WITHIN THE TEXT

Citations within the text should be based on the `natbib` package and include the authors' last names and year (with the "et al." construct for more than two authors). When the authors or the publication are included in the sentence, the citation should not be in parenthesis using (as in "See for more information."). Otherwise, the citation should be in parenthesis using (as in "Deep learning shows promise to make progress towards AI").

The corresponding references are to be listed in alphabetical order of authors, in the REFERENCES section. As to the format of the references themselves, any style is acceptable as long as it is used consistently.

8.2 FOOTNOTES

Indicate footnotes with a number¹ in the text. Place the footnotes at the bottom of the page on which they appear. Precede the footnote with a horizontal rule of 2 inches (12 picas).²

8.3 FIGURES

All artwork must be neat, clean, and legible. Lines should be dark enough for purposes of reproduction; art work should not be hand-drawn. The figure number and caption always appear after the figure. Place one line space before the figure caption, and one line space after the figure. The figure caption is lower case (except for first word and proper nouns); figures are numbered consecutively.

Make sure the figure caption does not get separated from the figure. Leave sufficient space to avoid splitting the figure and figure caption.

You may use color figures. However, it is best for the figure captions and the paper body to make sense if the paper is printed either in black/white or in color.

8.4 TABLES

All tables must be centered, neat, clean and legible. Do not use hand-drawn tables. The table number and title always appear before the table. See Table 1.

Place one line space before the table title, one line space after the table title, and one line space after the table. The table title must be lower case (except for first word and proper nouns); tables are numbered consecutively.

9 DEFAULT NOTATION

In an attempt to encourage standardized notation, we have included the notation file from the textbook, *Deep Learning* available at https://github.com/goodfeli/dlbook_

¹Sample of the first footnote

²Sample of the second footnote

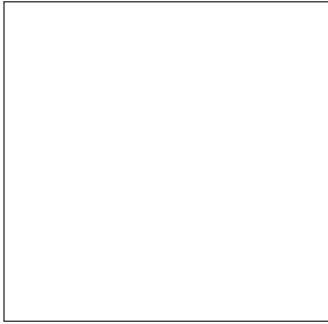


Figure 11: Sample figure caption.

Table 1: Sample table title

PART	DESCRIPTION
Dendrite	Input terminal
Axon	Output terminal
Soma	Cell body (contains cell nucleus)

notation/. Use of this style is not required and can be disabled by commenting out `math_commands.tex`.

Numbers and Arrays

a	A scalar (integer or real)
\boldsymbol{a}	A vector
\boldsymbol{A}	A matrix
\mathbf{A}	A tensor
\boldsymbol{I}_n	Identity matrix with n rows and n columns
\boldsymbol{I}	Identity matrix with dimensionality implied by context
$\boldsymbol{e}^{(i)}$	Standard basis vector $[0, \dots, 0, 1, 0, \dots, 0]$ with a 1 at position i
$\text{diag}(\boldsymbol{a})$	A square, diagonal matrix with diagonal entries given by \boldsymbol{a}
a	A scalar random variable
\boldsymbol{a}	A vector-valued random variable
\mathbf{A}	A matrix-valued random variable

Sets and Graphs

\mathbb{A}	A set
\mathbb{R}	The set of real numbers
$\{0, 1\}$	The set containing 0 and 1
$\{0, 1, \dots, n\}$	The set of all integers between 0 and n
$[a, b]$	The real interval including a and b
$(a, b]$	The real interval excluding a but including b
$\mathbb{A} \setminus \mathbb{B}$	Set subtraction, i.e., the set containing the elements of \mathbb{A} that are not in \mathbb{B}
\mathcal{G}	A graph
$Pa_{\mathcal{G}}(\mathbf{x}_i)$	The parents of \mathbf{x}_i in \mathcal{G}

Indexing

a_i	Element i of vector \mathbf{a} , with indexing starting at 1
\mathbf{a}_{-i}	All elements of vector \mathbf{a} except for element i
$A_{i,j}$	Element i, j of matrix \mathbf{A}
$\mathbf{A}_{i,:}$	Row i of matrix \mathbf{A}
$\mathbf{A}_{:,i}$	Column i of matrix \mathbf{A}
$\mathbf{A}_{i,j,k}$	Element (i, j, k) of a 3-D tensor \mathbf{A}
$\mathbf{A}_{:,:,i}$	2-D slice of a 3-D tensor
\mathbf{a}_i	Element i of the random vector \mathbf{a}

Calculus

$\frac{dy}{dx}$	Derivative of y with respect to x
$\frac{\partial y}{\partial x}$	Partial derivative of y with respect to x
$\nabla_{\mathbf{x}} y$	Gradient of y with respect to \mathbf{x}
$\nabla_{\mathbf{X}} y$	Matrix derivatives of y with respect to \mathbf{X}
$\nabla_{\mathbf{x}} \mathbf{y}$	Tensor containing derivatives of y with respect to \mathbf{X}
$\frac{\partial f}{\partial \mathbf{x}}$	Jacobian matrix $\mathbf{J} \in \mathbb{R}^{m \times n}$ of $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$
$\nabla_{\mathbf{x}}^2 f(\mathbf{x})$ or $\mathbf{H}(f)(\mathbf{x})$	The Hessian matrix of f at input point \mathbf{x}
$\int f(\mathbf{x}) d\mathbf{x}$	Definite integral over the entire domain of \mathbf{x}
$\int_{\mathbb{S}} f(\mathbf{x}) d\mathbf{x}$	Definite integral with respect to \mathbf{x} over the set \mathbb{S}

Probability and Information Theory

$P(a)$	A probability distribution over a discrete variable
$p(a)$	A probability distribution over a continuous variable, or over a variable whose type has not been specified
$a \sim P$	Random variable a has distribution P
$\mathbb{E}_{x \sim P}[f(x)]$ or $\mathbb{E}f(x)$	Expectation of $f(x)$ with respect to $P(x)$
$\text{Var}(f(x))$	Variance of $f(x)$ under $P(x)$
$\text{Cov}(f(x), g(x))$	Covariance of $f(x)$ and $g(x)$ under $P(x)$
$H(x)$	Shannon entropy of the random variable x
$D_{\text{KL}}(P\ Q)$	Kullback-Leibler divergence of P and Q
$\mathcal{N}(x; \mu, \Sigma)$	Gaussian distribution over x with mean μ and covariance Σ

Functions

$f : \mathbb{A} \rightarrow \mathbb{B}$	The function f with domain \mathbb{A} and range \mathbb{B}
$f \circ g$	Composition of the functions f and g
$f(x; \theta)$	A function of x parametrized by θ . (Sometimes we write $f(x)$ and omit the argument θ to lighten notation)
$\log x$	Natural logarithm of x
$\sigma(x)$	Logistic sigmoid, $\frac{1}{1 + \exp(-x)}$
$\zeta(x)$	Softplus, $\log(1 + \exp(x))$
$\ \mathbf{x}\ _p$	L^p norm of \mathbf{x}
$\ \mathbf{x}\ $	L^2 norm of \mathbf{x}
x^+	Positive part of x , i.e., $\max(0, x)$
$\mathbf{1}_{\text{condition}}$	is 1 if the condition is true, 0 otherwise

10 FINAL INSTRUCTIONS

Do not change any aspects of the formatting parameters in the style files. In particular, do not modify the width or length of the rectangle the text should fit into, and do not change font sizes (except perhaps in the REFERENCES section; see below). Please note that pages should be numbered.

11 PREPARING POSTSCRIPT OR PDF FILES

Please prepare PostScript or PDF files with paper size “US Letter”, and not, for example, “A4”. The `-t letter` option on `dvips` will produce US Letter files.

Consider directly generating PDF files using `pdflatex` (especially if you are a MiKTeX user). PDF figures must be substituted for EPS figures, however.

Otherwise, please generate your PostScript and PDF files with the following commands:

```
dvips mypaper.dvi -t letter -Ppdf -G0 -o mypaper.ps
ps2pdf mypaper.ps mypaper.pdf
```

11.1 MARGINS IN LATEX

Most of the margin problems come from figures positioned by hand using `\special` or other commands. We suggest using the command `\includegraphics` from the `graphicx` package.

Always specify the figure width as a multiple of the line width as in the example below using .eps graphics

```
\usepackage[dvips]{graphicx} ...
\includegraphics[width=0.8\linewidth]{myfile.eps}
```

or

```
\usepackage[pdftex]{graphicx} ...
\includegraphics[width=0.8\linewidth]{myfile.pdf}
```

for .pdf graphics. See section 4.4 in the graphics bundle documentation (<http://www.ctan.org/tex-archive/macros/latex/required/graphics/grfguide.ps>)

A number of width problems arise when LaTeX cannot properly hyphenate a line. Please give LaTeX hyphenation hints using the \- command.

AUTHOR CONTRIBUTIONS

If you'd like to, you may include a section for author contributions as is done in many journals. This is optional and at the discretion of the authors.

ACKNOWLEDGMENTS

Use unnumbered third level headings for the acknowledgments. All acknowledgments, including those to funding agencies, go at the end of the paper.

REFERENCES

- [1] Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, Erik Brynjolfsson, Shyamal Buch, Dallas Card, Rodrigo Castellon, Niladri Chatterji, Annie Chen, Kathleen Creel, Jared Quincy Davis, Dora Demszky, Chris Donahue, Moussa Doumbouya, Esin Durmus, Stefano Ermon, John Etchemendy, Kawin Ethayarajh, Li Fei-Fei, Chelsea Finn, Trevor Gale, Lauren Gillespie, Karan Goel, Noah Goodman, Shelby Grossman, Neel Guha, Tatsunori Hashimoto, Peter Henderson, John Hewitt, Daniel E. Ho, Jenny Hong, Kyle Hsu, Jing Huang, Thomas Icard, Saahil Jain, Dan Jurafsky, Pratyusha Kalluri, Siddharth Karamcheti, Geoff Keeling, Fereshte Khani, Omar Khattab, Pang Wei Koh, Mark Krass, Ranjay Krishna, Rohith Kuditipudi, Ananya Kumar, Faisal Ladhak, Mina Lee, Tony Lee, Jure Leskovec, Isabelle Levent, Xiang Lisa Li, Xuechen Li, Tengyu Ma, Ali Malik, Christopher D. Manning, Suvir Mirchandani, Eric Mitchell, Zanele Munyikwa, Suraj Nair, Avani Narayan, Deepak Narayanan, Ben Newman, Allen Nie, Juan Carlos Niebles, Hamed Nilforoshan, Julian Nyarko, Giray Ogut, Laurel Orr, Isabel Papadimitriou, Joon Sung Park, Chris Piech, Eva Portelance, Christopher Potts, Aditi Raghunathan, Rob Reich, Hongyu Ren, Frieda Rong, Yusuf Roohani, Camilo Ruiz, Jack Ryan, Christopher Ré, Dorsa Sadigh, Shiori Sagawa, Keshav Santhanam, Andy Shih, Krishnan Srinivasan, Alex Tamkin, Rohan Taori, Armin W. Thomas, Florian Tramèr, Rose E. Wang, William Wang, Bohan Wu, Jiajun Wu, Yuhuai Wu, Sang Michael Xie, Michihiro Yasunaga, Jiaxuan You, Matei Zaharia, Michael Zhang, Tianyi Zhang, Xikun Zhang, Yuhui Zhang, Lucia Zheng, Kaitlyn Zhou, and Percy Liang. On the opportunities and risks of foundation models. pages 3–4, 2022.
- [2] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
- [3] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale, 2021.

- [4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [5] Fan Liu, Tianshu Zhang, Wenwen Dai, Wenwen Cai, Xiaocong Zhou, and Delong Chen. Few-shot adaptation of multi-modal foundation models: A survey. pages 5–6, 2024.
- [6] Mateusz Wojcik. Continual learning: Methods and application. 2024.

A APPENDIX

You may include other additional sections here.