

الجمهورية الجزائرية الديمقراطية الشعبية
République algérienne démocratique et populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'enseignement supérieur et de la recherche scientifique
جامعة عين تموشنت بلحاج بوشعيب
Université –Ain Temouchent- Belhadj Bouchaib
Faculté des Sciences et de Technologie
Département D'électronique et des Télécommunications



Projet de Fin d'Etudes
Pour l'obtention du diplôme de Master en : Electronique
Domaine : Sciences et Technologie
Filière : Electronique
Spécialité : Instrumentation
Thème

Conception d'un système de contrôle de qualité

Présenté Par :

Mr Hamaida Habib

Devant le jury composé de :

Dr Badir Lahouaria

UAT.B.B (Ain Temouchent) Présidente

Dr Sekkal Mansouria

UAT.B.B (Ain Temouchent) Examinatrice

Mme Boutkhil Malika

UAT.B.B (Ain Temouchent) Encadrante

Dr Bentaieb Samia

UAT.B.B (Ain Temouchent) Co-encadrante

Année Universitaire 2021/2022

*Je dédie ce mémoire à **Mes chers parents**, pour tous leurs sacrifices, leur soutien et leurs prières tout au long de mes études, à mon grand frère **KHLED** et son fils **ADAM** et à **ma grande sœur** et ses enfants **ABD ELRAHMAN** et **YOUNES**, à tout ma famille et mes amis, À tous les enseignants qui m'ont enseigné durant ma carrière scolaire et universitaire. Je remercie tous mes camarades de la promotion **instrumentation 2022**, à toutes les personnes que j'ai rencontrées dans ma vie, que '**ALLAH** vous bénisse dans ta vie. .*

Habib

Remerciements

Avant tout, je tiens à remercier **ALLAH** le miséricordieux de pour m'avoir donné la santé et m'avoir accordé la volonté et la possibilité de réussir dans mes études, et de m'avoir donné courage d'accomplir ce travail.

Un grand Merci à mon encadrant Mme BOUTKHIL Malika pour son encadrement, et ses conseils précieux. Je tiens à remercier chaleureusement à ma co-encadrante Dr. BENTAIEB Samia pour le temps qu'elle a consacré et les conseils qu'elle m'a prodigués pour son encouragement sans son aide, ce travail n'aurait pas été réalisé. Avec son expérience dans la recherche et l'enseignement, avec ses conseils, j'ai pu découvrir le monde de la recherche scientifique dans le domaine de l'intelligence artificielle.

Je souhaite remercier Dr BADIR Lahouaria, qui m'a fait l'honneur de présider le jury. Je tiens aussi à remercier Dr SEKKAL Mansouria, pour avoir accepté d'examiner ce travail.

Je remercie également Mme Ikni Hind pour son aide précieuse et ses encouragements.

Enfin, je tiens à exprimer mes sincères remerciements pour **mes parents, ma famille**, mes collègues **Adel** et **Nasser** pour les moments consacrés ensemble à la réalisation de nos projets. Merci à tous ceux qui ont participé à la réalisation de ce projet, tant au niveau humain qu'au niveau scientifique.

s

Table des figures

1.1	Représentation d'un pixel dans une image	8
1.2	Les différents types d'image	8
1.3	Représentation d'image couleurs par trois matrices	9
1.4	Une image avec différentes résolutions	10
1.5	Une image avec différents types de bruit	12
1.6	Une image filtrée par un filtre median	13
1.7	Exemple d'une image recadrée	14
1.8	Une image redimensionnée avec différents types d'interpolationt	15
1.9	Illustration de la notion de FPS	17
1.10	Comparatif des différentes résolutions	17
2.1	Type de Machine Learning	21
2.2	Diagramme illustrant les concepts de base de l'intelligence artificielle, de l'apprentissage automatique et de l'apprentissage profond.[Messaoudi, 2020]	22
2.3	Architecture d'un neurone biologique	23
2.4	Illustration de réseau de neurones	24
2.5	Architecture d'un neurone artificiel (perceptron)	25
2.6	Les fonctions d'activations	27
2.7	Exemple du réseau de neurones convolutifs CNN 2D(classification d'une bouteille)	27
2.8	Couche de convolution 2D	28
2.9	Max Pooling	29
2.10	Exemple d'une couche entièrement connectée	30
2.11	la différence entre original networks et dropout networks	30

2.12	Exemple de flattening 2D	31
3.1	Architecture du modèle de contrôle de qualité	38
3.2	Exemple de bouteilles défectueuses et non défectueuses	40
3.3	La matrice de confusion de la qualité	42
3.4	Le taux d'apprentissage e l'erreur en fonction du nombre d'époques .	43
3.5	Exemple de bouteilles défectueuses et non défectueuses	43

Liste des tableaux

3.1	Matrice de confusion d'un problème à deux classes	41
-----	---	----

Listings

3.1 Répartition de la base de données	39
3.2 Hyper paramètres et compilation du modèle	40

Abréviations

AI : **A**rtificielle **I**ntelligence

ANN : **A**rtificial **N**eural **N**etworks

BP : **B**ack **P**ropagation

CNN : **C**onvolution **N**eural **N**etworks

CPU : **C**entral **P**rocessing **U**nit DL : **D**eep **L**earning

DPI : **D**ots **P**er **I**ch

FPS : **F**rames **P**er **S**econd

GIF : **G**raphics **I**nterchange **F**ormat

GPIO : **G**eneral **P**urpose **I**nput/**O**utput

GPU : **G**raphics **P**rocessing **U**nit

HD : **H**igh **D**efinition

HDMI : **M**ultimedia **I**nterface

JPEG : **J**oint **P**hotographic **E**xperts **G**roup

LCD : **L**iquid **C**rystal **D**isplay

ML : **M**achine **L**earning

OS : **O**perating **S**ystem

PCA : **P**rinical **A**nalysis

ReLU : **R**ectified **L**inear **U**nits

ROI : **R**egionl **O**of **I**nterest

RNN : **R**ecurrent **N**eural **N**etworks

SBC : **S**ingle **B**oard **C**omputer PNG : **P**ortable **N**etwork **G**raphics

SD : **S**tandard **D**efinition

SGD : **S**tochastic **G**radient **D**escent SOC : **S**ystem **O**n **C**hip

SVM : **S**upport **V**ector **M**achine

TIFF : **T**agged **I**mage **F**ile **F**ormat
USB : **U**niversal **S**erial **B**us
VGA : **V**ideo **G**raphics **A**rray
VNC : **V**irtual **N**etwork **C**omputing

Résumé

Les fabricants de boissons sont obligés de contrôler la qualité de leurs produits, ce qui les oblige à employer plus de travailleurs dans les lignes de production pour contrôler la qualité du produit. Cela conduit à l'augmentation du taux d'erreurs humaines à cause de la fatigue et de l'épuisement entraînant des pertes financières pour l'entreprise.

Dans ce travail, nous proposons un système automatique de contrôle, qui utilise la vision par ordinateur afin de réduire le risque de perte financière et d'augmenter la productivité et la diversification des produits. Le principe de fonctionnement de notre système est le suivant : lorsque la bouteille se trouve dans la zone de détection du capteur infrarouge sur le convoyeur, les bouteilles sont détectées par le capteur. La caméra prend une image de la bouteille en temps réel. Ce traitement utilise une classification basée sur l'apprentissage profond avec un réseau de neurones convolutifs entraîné pour détecter les produits défectueux.

Le système proposé peut remplacer le contrôle de qualité classique basé sur les inspections manuelles. Ce système présente l'avantage d'une vérification rapide ainsi que plusieurs types de vérification en même temps. Il permet d'augmenter la productivité, d'améliorer la qualité de l'inspection, la variété des produits et la rentabilité.

Mots clés : intelligence artificielle, apprentissage profond , contrôle de la qualité ,réseau neuronal convolutif (CNN).

ملخص

يضطر مصنو المزيد من العمال في خطوط الإنتاج للتحكم في جودة المنتج. وهذا يؤدي إلى زيادة في معدل الخطأ البشري بسبب التعب والإرهاق مما يؤدي إلى خسائر مالية للشركة. في هذا العمل، نقترح نظام تحكم تلقائي، يستخدم رؤية الكمبيوتر لتقليل مخاطر الخسارة المالية وزيادة الإنتاجية وتنويع المنتجات. مبدأ التشغيل في نظامنا هو كما يلي :

عندما تكون الزجاجات في منطقة الكشف عن مستشعر الأشعة تحت الحمراء على الناقل، يتم اكتشاف الزجاجات بواسطة المستشعر. تأخذ الكاميرا صورة في الوقت الفعلي للزجاجة. يستخدم هذا العلاج تصنيفاً قائماً على التعلم العميق مع شبكة من الخلايا العصبية التلافيفية المدربة على اكتشاف المنتجات المعيبة. ويمكن أن يحل النظام المقترح محل مراقبة الجودة التقليدية القائمة على عمليات التفتيش اليدوية. ويتميز هذا النظام بالتحقق السريع وكذلك بعدة أنواع من التحقق في الوقت نفسه. إنه يزيد الإنتاجية ويحسن جودة التفتيش وتنوع المنتجات والربحية المشروبات إلى التحكم في جودة منتجاتهم، مما يجبرهم على توظيف

الكلمات الرئيسية : الذكاء الاصطناعي، التعلم العميق، مراقبة الجودة، الشبكة العصبية التلافيفية ، ، CNN .

Abstract

Beverage manufacturers are required to control the quality of their products, which forces them to employ more workers in the production lines to control the quality of the product. This leads to an increase in the rate of human error due to fatigue and exhaustion resulting in financial losses for the company. In this work, we propose an automatic control system, which uses computer vision to reduce the risk of financial loss and increase productivity and product diversification. The working principle of our system is as follows : when the bottle is in the detection area of the infrared sensor on the conveyor, the bottles are detected by the sensor. The camera takes an image of the bottle in real time. This processing uses deep learning-based classification with a trained convolutional neural network to detect defective products. The proposed system can replace conventional quality control based on manual inspections. This system has the advantage of fast verification as well as multiple types of verification at the same time. It can increase productivity, improve inspection quality, product variety and profitability.

Keywords : artificial intelligence, deep learning, quality control ,convolutional neural network (CNN).

Introduction générale

Contexte

De nos jours, le marché de l'industrie manufacturière tend à se caractériser de plus en plus par une concurrence intense et féroce, poussant toutes les usines à rechercher les méthodes et les outils nécessaires pour améliorer la réputation de leur produit. Cependant, le contrôle de la qualité des produits à toutes les étapes de la fabrication jusqu'à l'étape de la distribution est souvent le principal problème industriel. Le système de vision artificielle a attiré l'attention de la recherche en industrie depuis les années 1970. A la fin des années 1980 et au début des années 1990, les avancés ont augmenté de manière spectaculaire, en grande partie grâce à des avancées technologiques significatives dans la fabrication de matériel de traitement permettant une numérisation plus rapide des images.

Le système de vision artificielle est généralement désigné comme le système qui extrait les caractéristiques souhaitées des images numériques. Les images d'entrée capturées sont l'objectif principal de ce système. L'application du système de vision dans les industries afin d'automatiser le processus de fabrication est considérée comme une inspection visuelle automatisée lorsque des tentatives sont faites pour inspecter, contrôler les produits et reconnaître les défauts en utilisant uniquement des images des produits . En fait, les inspecteurs humains sont plus lents et leur efficacité est affectée par leur état de santé, leur épuisement ou d'autres défauts humains.

Objective

Dans ce travail, nous présentons le principe de base de l'inspection visuelle artificielle des bouteilles en présentant différentes méthodes et techniques pour l'automatisation de la mise en bouteille et les algorithmes les plus récents ont été présentés. Nous nous sommes concentrés sur la détection des produits (bouteille) défectueux qui sont inspectés automatiquement.

Structure

La partie restante du travail est organisée en 3 chapitres. Dans le chapitre 1, nous présentons d'abord les concepts de base nécessaires à la compréhension des différentes méthodes de traitement d'images. Le chapitre 2 donne un aperçu détaillé sur l'intelligence artificielle, ses types, particulièrement le deep learning. Le chapitre 3 couvre le méthodologie que nous avons suivie pour mener le contrôle de qualité des bouteilles, y compris la collecte de données, les outils utilisés pour mettre en œuvre le modèle et les techniques d'évaluation sont discutées. Finalement, une conclusion générale est présentée.

Chapitre 1

Traitement d'Images

Sommaire

1.1	Introduction	7
1.2	Préliminaires	7
1.2.1	Image numérique	7
1.2.2	Les types d'images	8
1.2.3	La résolution d'une image	9
1.2.4	Les formats d'images	9
1.3	Le prétraitement des images	11
1.3.1	Élimination du bruit	12
1.3.2	Recadrage d'une image	13
1.3.3	Redimensionnement d'une image	14
1.4	La vidéo	16
1.4.1	Définition	16
1.4.2	Nombre d'images par seconde	16
1.4.3	La résolution	17
1.4.4	Domaine d'utilisation de la vidéo	18
1.5	Conclusion	18

1.1 Introduction

Les êtres humains dépendent principalement de leur système de vision pour donner un sens au monde qui les entoure. Ils ne regardent pas seulement les choses pour les identifier et les classer, mais aussi pour trouver les différences et obtenir une "impression" générale d'une scène en un coup d'œil rapide.

1.2 Préliminaires

Le traitement d'images désigne toute forme de traitement du signal dont l'entrée est une image, telle qu'une photographie ou une image vidéo, la sortie du traitement d'images pouvant être soit une image, soit un ensemble de caractéristiques ou de paramètres liés à l'image [Petrou and Petrou, 2010].

Le traitement d'images est aussi une méthode permettant de convertir une image sous forme numérique et d'effectuer certaines opérations sur celle-ci, afin d'obtenir une image améliorée ou d'en extraire des informations utiles.

Les techniques de traitement numérique permettent de manipuler les images numériques à l'aide d'ordinateurs. Les données brutes provenant des capteurs d'imagerie contiennent généralement des imperfections. Pour surmonter ces défauts, elles doivent subir différentes phases de traitement [Castleman, 1996]. Il existe des principes de traitement d'image de base que nous devons connaître avant d'entamer ce projet. Ce chapitre donne une vue d'ensemble du contexte théorique des méthodes utilisées dans le traitement des images.

1.2.1 Image numérique

Une image peut être définie comme une fonction bidimensionnelle, $I(x, y)$, où x et y sont des coordonnées spatiales, et l'amplitude de I à n'importe quelle paire de coordonnées (x, y) est appelée intensité de l'image à ce point [Castleman, 1996].

Une image numérique est composée d'un ensemble de points ou d'éléments d'image, généralement appelés pixels, stockés dans un tableau de chiffres. Les images sont des données spatiales indexées par deux coordonnées spatiales. Le système de coordonnées d'une image I est défini comme illustré dans la figure 1.1. L'image est représentée par $I(x, y)$, où x est la position horizontale du pixel et y la position verticale. Pour la petite image de la figure 1.1, $I(0, 0) = 10$, $I(3, 1) = 95$ et $I(2, 3) = 19$.

0	0	255	11
200	255	25	173
85	10	5	0
120	18	185	19

Pixel

FIGURE 1.1 – Représentation d'un pixel dans une image

1.2.2 Les types d'images

Une image binaire comme son nom l'indique, ne contient que deux éléments de pixel, à savoir 0 et 1, où 0 correspond au noir et 1 au blanc .

Une image en niveaux de gris est simplement une image en noir et blanc où elle est représentée par une seule composante d'intensité. Elle possède 256 teintes de gris, la valeur zéro étant le noir et la valeur 255 le blanc. Le nombre 256 est lié à la quantification de l'image. Chaque entier représentant un niveau de gris est codé sur 8 bits. Il est donc compris entre 0 et $2^8 - 1 = 255$.

Une image RGB est une image à trois composantes de couleur où les composantes sont le rouge (R), le vert (G) et le bleu (B) (voir la figure 1.3). L'intensité de la couleur de chaque composante est représentée par un nombre de 8 bits. la figure 1.2 montre un exemple des types d'images mentionnés.



FIGURE 1.2 – Les différents types d'image

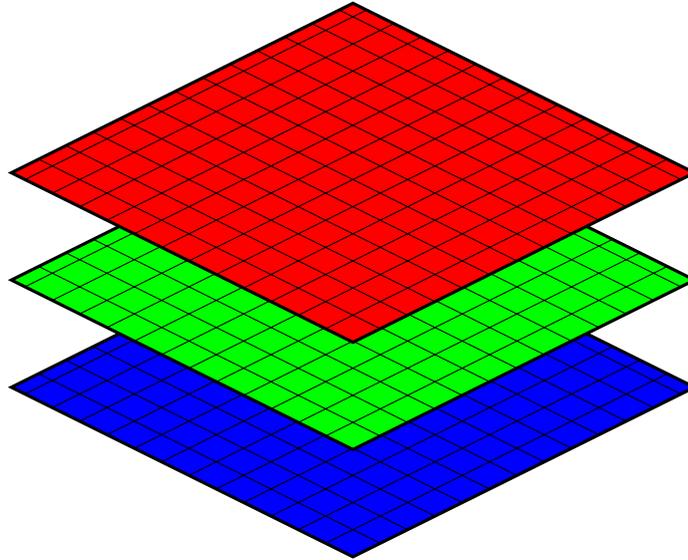


FIGURE 1.3 – Représentation d’image couleurs par trois matrices

Une image couleur RGB peut être convertie en image en niveaux de gris en utilisant différentes pondérations des composantes rouge, verte et bleue. L’équation peut être la suivante

$$Gray(i, j) = 0.40 \times R(i, j) + 0.60 \times G(i, j) + 0.30 \times B(i, j) \quad (1.1)$$

1.2.3 La résolution d’une image

C’est le nombre de points contenu dans une longueur donnée (en pouce). Elle est exprimée en Dots Per Inch (DPI) ; un inch mesure 2.54 cm. La résolution permet ainsi d’établir le rapport entre la définition en pixels d’une image et la dimension réelle de sa représentation sur un support physique (affichage écran, impression papier...). La figure 1.4 montre la même image avec différentes résolutions.

1.2.4 Les formats d’images

Il existe 4 formats principaux pour stocker les images :

1. TIFF

TIFF est l’abréviation de Tagged Image File Format. Les images TIFF créent des fichiers de très grande taille. Les images TIFF ne sont pas compressées et contiennent donc beaucoup de données d’image détaillées (c’est pourquoi les

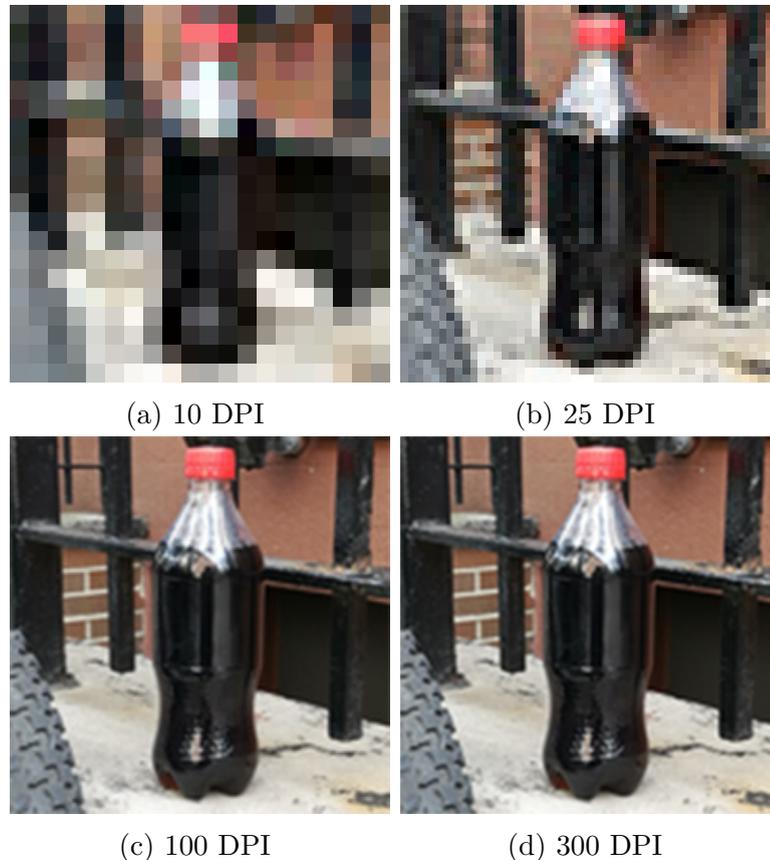


FIGURE 1.4 – Une image avec différentes résolutions

fichiers sont si volumineux). Les TIFF sont également extrêmement flexibles en termes de couleur (ils peuvent être en niveaux de gris, ou CMYK pour l'impression, ou RGB pour le web).

2. JPEG

JPEG est l'abréviation de Joint Photographic Experts Group, qui a créé cette norme pour ce type de format d'image. Les fichiers JPEG sont des images qui ont été compressées pour stocker beaucoup d'informations dans un fichier de petite taille. La plupart des appareils photo numériques stockent les photos au format JPEG, car il est alors possible de prendre plus de photos sur une carte d'appareil photo qu'avec d'autres formats.

Un JPEG est compressé de manière à perdre une partie des détails de l'image au cours de la compression, afin de réduire la taille du fichier (on parle alors de compression "avec perte").

Les fichiers JPEG sont généralement utilisés pour les photographies sur le Web, car ils créent un petit fichier qui se charge facilement sur une page Web et qui est de bonne qualité.

Les fichiers JPEG sont mauvais pour les dessins au trait, les logos ou les graphiques, car la compression leur donne un aspect "bitmappy" (lignes irrégulières au lieu de lignes droites).

3. GIF

GIF est l'abréviation de Graphic Interchange Format. Ce format compresse les images mais, à la différence du JPEG, la compression est sans perte (aucun détail n'est perdu dans la compression, mais le fichier ne peut pas être rendu aussi petit qu'un JPEG).

Les GIF ont également une gamme de couleurs extrêmement limitée, adaptée au Web mais pas à l'impression. Ce format n'est jamais utilisé pour la photographie, en raison du nombre limité de couleurs. Les GIF peuvent également être utilisés pour les animations.

4. PNG

Le PNG est l'abréviation de Portable Network Graphics. Il a été créé en tant que format ouvert pour remplacer le GIF, car le brevet du GIF était détenu par une société et personne d'autre ne voulait payer les frais de licence. Il permet également une gamme complète de couleurs et une meilleure compression.

Il est utilisé presque exclusivement pour les images Web, jamais pour les images imprimées. Pour les photographies, le PNG n'est pas aussi bon que le JPEG, car il crée un fichier plus volumineux. Mais pour les images contenant du texte ou des dessins au trait, le PNG est meilleur, car les images ont l'air moins "bitmappy".

1.3 Le prétraitement des images

Le prétraitement est utilisé pour supprimer le bruit et éliminer les informations non pertinentes et visuellement inutiles. Le bruit est une information indésirable qui peut résulter du processus d'acquisition de l'image. D'autres étapes de prétraitement peuvent inclure la recherche de régions intéressantes pour un traitement ultérieur.

1.3.1 Élimination du bruit

Le bruit est toute information indésirable qui vient contaminer une image. Il apparaît dans l'image à partir d'une variété de sources. Lors du processus d'acquisition d'images numériques, l'image optique est convertit en un signal électrique continu qui est ensuite échantillonné. Ceci est le principal processus par lequel le bruit apparaît dans les images numériques. D'autres facteurs tels que le mauvais fonctionnement du capteur, les défauts de la procédure de transmission des données et les interférences électroniques peuvent réduire de manière significative les images [Castleman, 1996].

Tout ceci conduit à l'ajout d'une valeur aléatoire à la luminosité exacte d'un pixel donné. Dans une image, le bruit peut être modélisé avec l'une des distributions suivantes :

- Distribution Gaussienne
- Distribution uniforme
- Distribution du sel et du poivre.

La figure 1.6 montre un exemple des types d'images mentionnés.

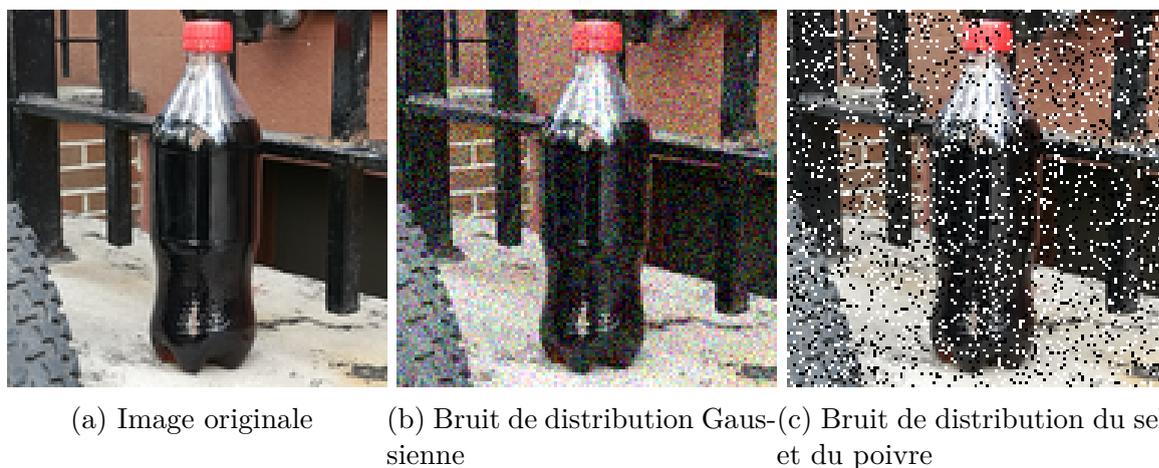


FIGURE 1.5 – Une image avec différents types de bruit

Afin de supprimer différents types de bruit dans les images numériques et effectuer un certain type d'amélioration de l'image, un filtrage spatial est généralement utilisé. Les trois types de filtres sont :

- Les filtres moyens
- Filtres médians
- Filtres de rehaussement

Les filtres moyens et médians sont principalement utilisés pour masquer ou supprimer le bruit, mais ils peuvent également être utilisés pour des applications spéciales. Par exemple, un filtre moyen donne un aspect plus "doux" à une image. Le filtre de rehaussement fait ressortir les bords et les détails de l'image. Les filtres spatiaux sont implémentés avec des masques de convolution. Comme l'opération du masque de convolution fournit un résultat qui est la somme pondérée des valeurs d'un pixel et de ses voisins, on l'appelle un filtre linéaire [Andrews et al., 2013].



(a) Image originale (b) Image bruitée avec un de(c) Bruit de distribution du sel et du poivre
distribution du sel et du poivre

FIGURE 1.6 – Une image filtrée par un filtre median

1.3.2 Recadrage d'une image

Souvent, pour l'analyse d'images, nous voulons étudier de plus près une zone spécifique de l'image appelée Region Of Interest (ROI). Pour ce faire, nous avons besoin d'une opération qui modifie les coordonnées spatiales de l'image, appelée recadrage.

Le processus de recadrage d'une image consiste à sélectionner une petite partie de l'image, une sous-image et à la couper du reste de l'image. La figure 1.7 montre un exemple d'une image recadrée aux points $P_1(15, 15)$, $P_2(150, 100)$.

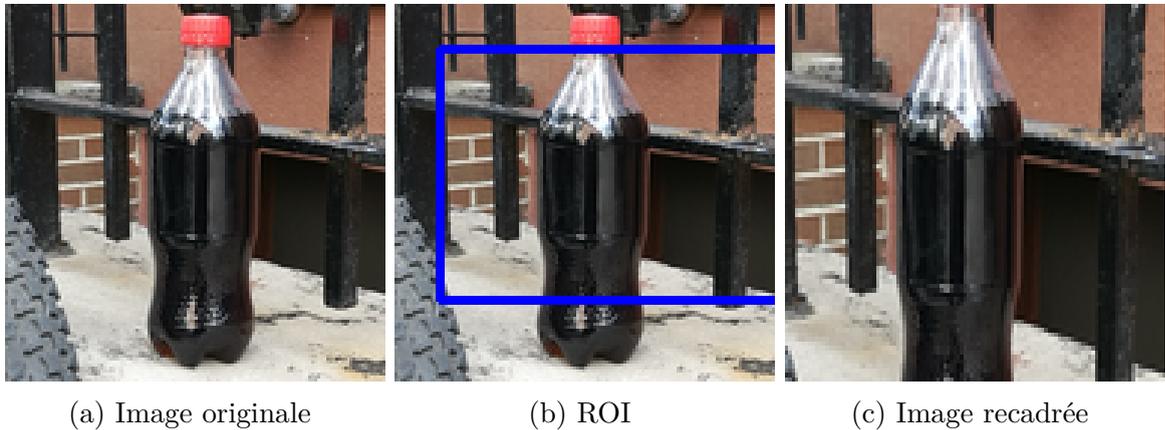


FIGURE 1.7 – Exemple d’une image recadrée

1.3.3 Redimensionnement d’une image

Le redimensionnement des images fait référence à la mise à l’échelle d’une image ou à la déformation de la grille de pixels originale à une autre. Cette opération est utile dans de nombreuses applications de traitement d’images et d’apprentissage automatique. Elle permet de réduire le nombre de pixels d’une image. Le redimensionnement permet également de zoomer dans les images pour répondre aux exigences de l’application utilisée.

Le redimensionnement se fait par une opération d’interpolation d’image bidimensionnelle où différentes techniques sont employées.

L’image la plus grossière obtenue provient de l’interpolation du plus proche voisin. Cette méthode d’interpolation se contente de regarder la valeur de la fonction la plus proche d’un point donné et lui attribue la même valeur, ce qui conduit à une image très pixélisée. L’interpolation bilinéaire rapproche la fonction en utilisant une approximation linéaire, ce qui permet d’obtenir une meilleure image. L’interpolation bicubique rapproche la fonction en utilisant l’approximation cubique, ce qui permet d’obtenir une image encore meilleur [Thévenaz et al., 2000].

La figure 1.8 montre un exemple d’une image avant et après le redimensionnement.



(a) Image originale (256×256) (b) (Image réduite (64×64))



(c) Interpolation du plus proche voisin (d) Interpolation bilinéaire



(e) Interpolation cubique

FIGURE 1.8 – Une image redimensionnée avec différents types d'interpolation

1.4 La vidéo

L'une des plus grandes avancées technologiques de cette ère a été l'émergence de la vidéo numérique comme élément fondamental de la vie quotidienne. Une séquence vidéo contient plus d'informations visuelles qu'une seule image. Cela est dû en partie au fait que le mouvement est capturé dans une séquence d'images, plutôt que dans une seule image.

1.4.1 Définition

La vidéo est la technologie qui permet de capturer des images en mouvement par voie électronique. Ces images en mouvement sont en réalité juste une série d'images fixes qui changent si vite qu'on dirait que l'image bouge. La fluidité (vitesse) d'une vidéo est caractérisée par le nombre d'images par secondes (en anglais Frames per second (FPS)).

Une image numérique dans un flux vidéo est représentée par une matrice de valeurs, où chaque valeur est une fonction de l'information entourant le point correspondant dans l'image. Un élément unique dans une matrice d'image comprend des informations pour toutes les composantes de couleur.

Pour stocker ou transmettre une vidéo, une grande capacité de stockage et un taux de transfert rapide sont nécessaires. L'un des facteurs les plus importants de la compression vidéo est le nombre d'images par seconde et la résolution.

1.4.2 Nombre d'images par seconde

Le nombre d'images par seconde (FPS) est la vitesse à laquelle des images successives, appelées trames, apparaissent sur un écran et forment des images en mouvement. Le contenu vidéo que nous consommons quotidiennement n'est pas réellement en mouvement. En fait, il s'agit d'images fixes qui s'enchaînent les unes après les autres. Si une vidéo est tournée à 24 images par seconde, cela signifie que 24 images individuelles sont lues en une seconde (voir la figure 1.9). Elles changent à un rythme différent selon les supports, en fonction de nombreux autres facteurs. [FPS,]

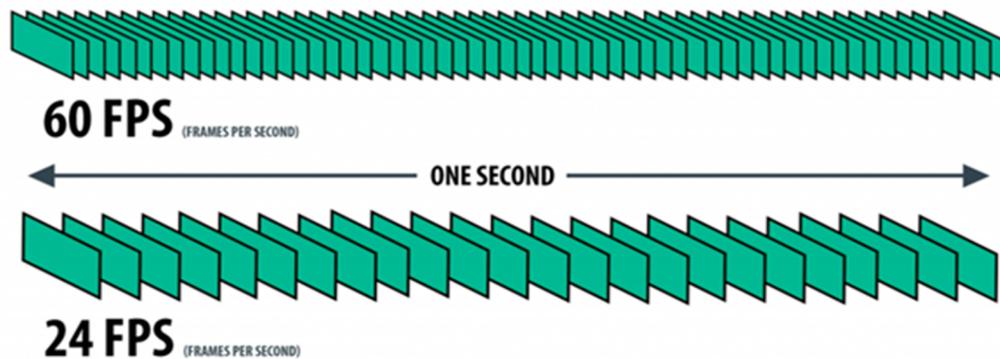


FIGURE 1.9 – Illustration de la notion de FPS

1.4.3 La résolution

La résolution d'un dispositif d'affichage est définie comme le nombre de pixels qui peuvent être affichés. Elle est généralement exprimée par des nombres pour l'horizontale et la verticale : 640 par 480 signifie 640 pixels de largeur, par 480 pixels de hauteur donnant 307.200 pixels. Il existe de nombreux formats de résolution tels que le SD, HD, le Full HD, etc. La figure 1.10 donne un comparatif entre les différentes résolutions.

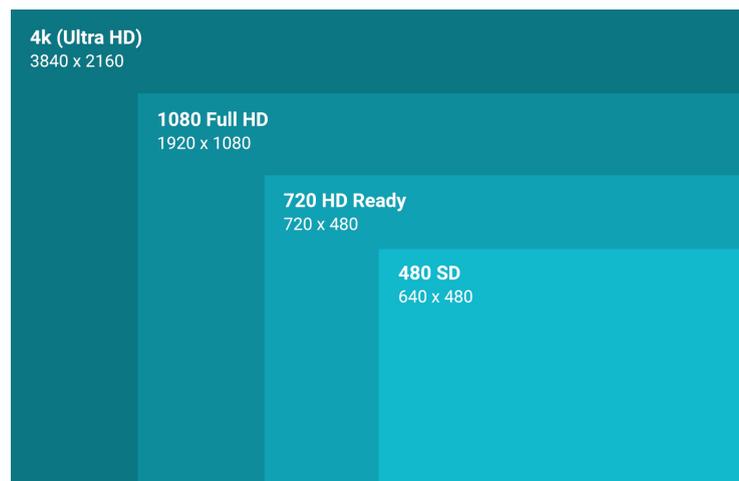


FIGURE 1.10 – Comparatif des différentes résolutions

1.4.4 Domaine d'utilisation de la vidéo

Au cours des deux dernières décennies, on a assisté à un nombre croissant d'applications de la vidéo dans de nombreux domaines, notamment les communications, l'éducation, la médecine et le divertissement. Ces applications vidéo améliorent la communication, favorisent une compréhension plus rapide d'idées complexes et offrent des capacités d'accès accrues à l'information.

Différentes tâches de vidéo-surveillance telles que la détection de mouvement, le contrôle des scènes intérieures et extérieures comme les aéroports, gares, parkings, autoroutes, magasins, etc. sont de plus en plus demandées. le contrôle de qualité en utilisant la vision par ordinateur est un domaine en plein essor.

1.5 Conclusion

Ce chapitre a introduit les concepts de base nécessaires à la compréhension des différentes méthodes de traitement d'images. Il s'est également penché sur les caractéristiques des vidéos. Les chapitres suivants traiteront des méthodes de contrôle de qualité des bouteilles basées sur le traitement d'images.

Chapitre 2

Intelligence Artificielle

Sommaire

2.1	Introduction	20
2.2	Intelligence artificielle	20
2.2.1	Machine learning	20
2.2.2	Deep Learning	22
2.2.3	Les fonctions d'activation	25
2.2.4	Apprentissage d'un réseau de neurones convolutifs (Neuron learning) :	30
2.3	Conclusion	33

2.1 Introduction

L'inspection automatique des défauts a fait l'objet d'une attention accrue dans le cadre du contrôle de la qualité des produits industriels. La détection des défauts est généralement effectuée dans le cadre de scénarios industriels complexes, ce qui constitue un problème intéressant mais difficile. Les méthodes traditionnelles sont basées sur le traitement d'images ou sur des techniques d'apprentissage automatique superficiel, mais elles ne peuvent détecter les défauts que dans des conditions de détection spécifiques, telles que des contours de défauts évidents avec un fort contraste et un faible bruit, à certaines échelles ou sous un éclairage spécifique. Dans ce chapitre, nous abordons le sujet de l'intelligence artificielle qui sera la pierre angulaire de notre travail.

2.2 Intelligence artificielle

L'intelligence artificielle¹ (IA) fait référence à la simulation de l'intelligence humaine dans des machines qui sont programmées pour penser comme les humains et imiter leurs actions. Le terme peut également s'appliquer à toute machine qui présente des caractéristiques associées à l'esprit humain, comme l'apprentissage et la résolution de problèmes.

Pour atteindre l'intelligence artificielle, un système d'IA utilise principalement l'apprentissage automatique ou le machine learning et d'autres techniques d'analyse de données.

2.2.1 Machine learning

L'apprentissage automatique est un sous-domaine de l'intelligence artificielle. L'idée derrière l'apprentissage automatique est de créer une application qui apprend de son expérience. Pour que cette application puisse apprendre quoi que ce soit, un ensemble de données doit être fourni. Il n'est donc pas nécessaire de créer explicitement des algorithmes pour prédire le résultat de nouveaux échantillons de données. C'est l'application d'apprentissage automatique qui s'en charge. L'apprentissage automatique peut être appliqué soit dans une approche supervisée, soit dans une approche non supervisée ou une approche semi-supervisée [Mitchell and Mitchell,

1. <https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp>

1997].

La figure 2.1 donne une illustration des types du Machine Learning.

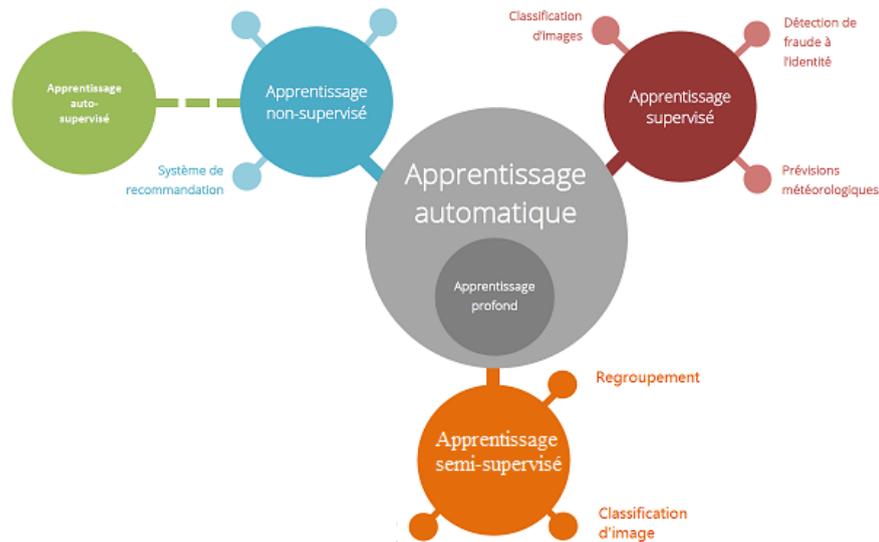


FIGURE 2.1 – Type de Machine Learning

2.2.1.1 Apprentissage supervisé

L'apprentissage supervisé est couramment utilisé pour les problèmes de classification ou de régression. Lorsque les étiquettes sont représentées sous forme de catégories, telles que cap et no cap, le problème est un problème de classification. Lorsque les étiquettes sont représentées par une valeur numérique, comme un prix, des hauteurs en centimètres, etc., le problème est un problème de régression [Bishop and Nasrabadi, 2006].

2.2.1.2 Apprentissage non-supervisé

Dans l'approche de l'apprentissage non supervisé, l'ensemble de données se compose uniquement de caractéristiques d'entrée. Les étiquettes sont absentes et les algorithmes d'apprentissage automatique ne peuvent pas exploiter le fait de connaître les véritables étiquettes pendant le processus d'apprentissage. Les approches d'apprentissage non supervisé recherchent une structure dans l'ensemble de données [Bishop and Nasrabadi, 2006].

2.2.1.3 Apprentissage semi-supervisé

L'apprentissage semi-supervisé est similaire à l'apprentissage supervisé, mais il utilise des données étiquetées et non étiquetées. Les données étiquetées sont essentiellement des informations comportant des balises significatives permettant à l'algorithme de comprendre les données, tandis que les données non étiquetées ne comportent pas ces informations. En utilisant cette combinaison, les algorithmes d'apprentissage automatique peuvent apprendre à étiqueter les données non étiquetées [Bishop and Nasrabadi, 2006].

2.2.2 Deep Learning

Le Deep learning est l'une des principales technologies de machine learning et d'intelligence artificielle (voir la figure 2.2).

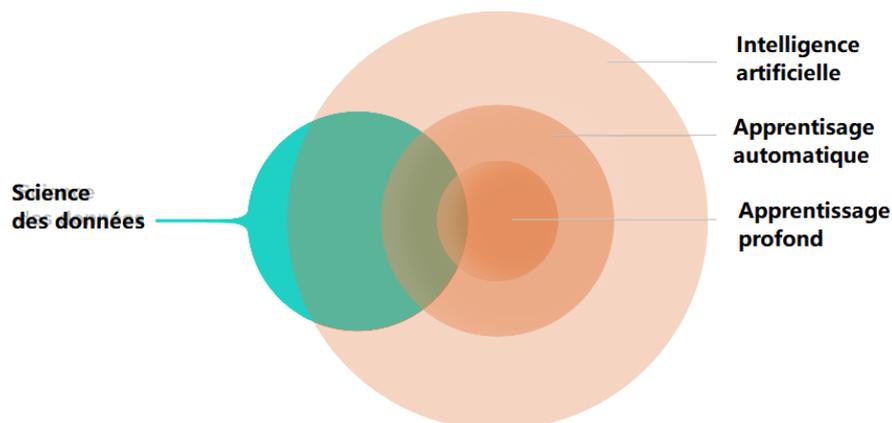


FIGURE 2.2 – Diagramme illustrant les concepts de base de l'intelligence artificielle, de l'apprentissage automatique et de l'apprentissage profond.[Messaoudi, 2020]

Un modèle de deep learning est conçu pour analyser continuellement les données avec une structure logique similaire à la façon dont un humain tirerait des conclusions. Pour effectuer cette analyse, les applications d'apprentissage profond utilisent une structure d'algorithmes en couches appelée réseau neuronal artificiel. La conception d'un réseau neuronal artificiel s'inspire du réseau biologique de neurones du cerveau humain, ce qui permet d'obtenir un système d'apprentissage bien

plus performant que les modèles d'apprentissage automatique standard.

Pour des cas d'utilisation plus pratiques, imaginez une application de reconnaissance d'images capable d'identifier un type de fleur ou une espèce d'oiseau à partir d'une photo. Cette classification d'images est alimentée par un réseau neuronal profond. L'apprentissage profond guide également la reconnaissance et la traduction de la parole et conduit littéralement les voitures autonomes.

2.2.2.1 Le cerveau humain et les réseaux neuronaux artificiels (Artificial Neural Networks)

Les scientifiques et les chercheurs tentent de reproduire l'activité du cerveau humain. La figure reffig :neurone illustre les dendrites, le corps cellulaire ou soma, l'axone et les terminaisons axonales d'un neurone biologique. La réception des signaux provenant d'autres neurones se fait par l'intermédiaire des dendrites. Le soma détermine si un signal, appelé impulsion nerveuse, sera généré ou non en fonction des signaux qu'il reçoit. L'axone va transporter cet impulsion nerveuse jusqu'aux terminaisons axonales, où il sera transmis à d'autres neurones en aval par les synapses.

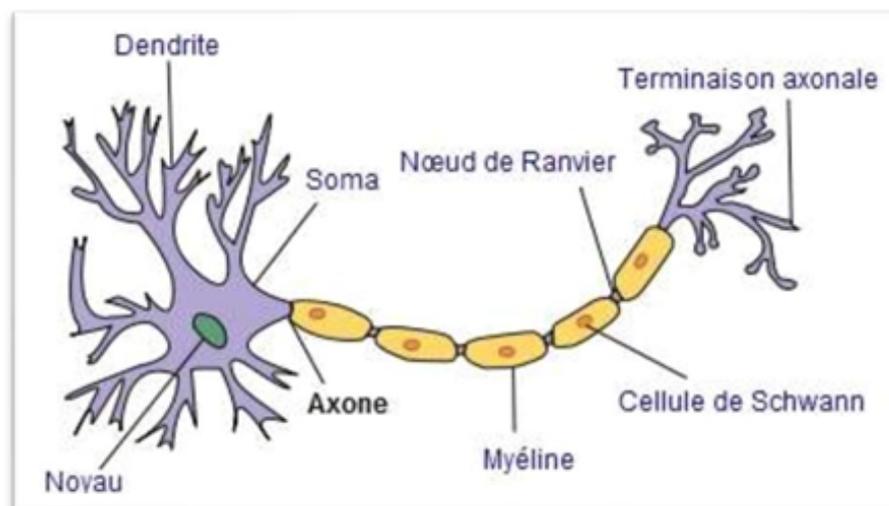


FIGURE 2.3 – Architecture d'un neurone biologique

Artificial Neural Networks (ANN) est un mécanisme d'apprentissage automatique très puissant qui imite essentiellement la façon dont le cerveau humain apprend. Le cerveau reçoit le stimulus du monde extérieur, effectue le traitement de

l'entrée, puis génère la sortie. Lorsque la tâche se complique, de multiples neurones forment un réseau complexe, transmettant des informations entre, comme le montre dans la figure 2.3.

2.2.2.2 L'architecture des réseaux neuronaux

Comme le montre la figure 2.4, les réseaux neuronaux comportent principalement 3 couches :

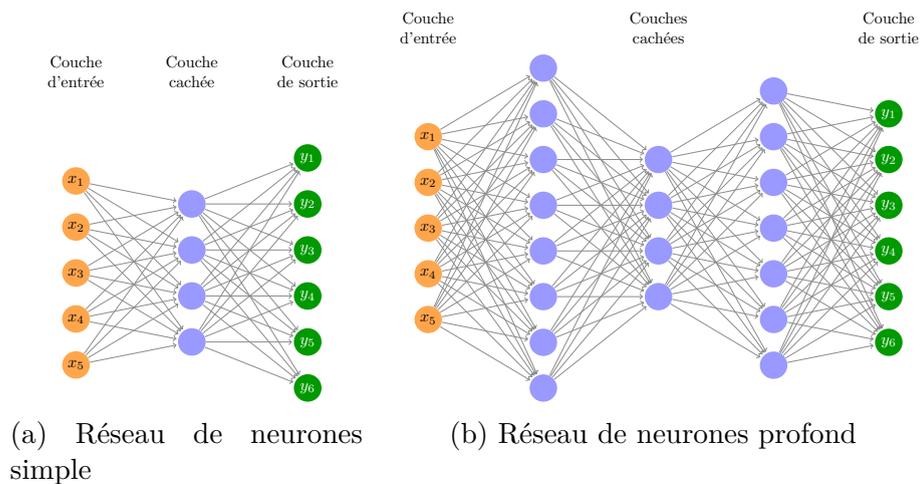


FIGURE 2.4 – Illustration de réseau de neurones

- Input Layer (couche d'entrée)
- Hidden Layers (couches cachées)
- Output Layer (couche de sortie)

Input Layer : La couche d'entrée contient les neurones pour l'entrée des caractéristiques. Un biais est également ajouté à la couche d'entrée en plus des caractéristiques.

Hidden Layers : Les couches cachées sont les couches intermédiaires entre les couches d'entrée et de sortie. Il peut y avoir un nombre quelconque de couches cachées. Les réseaux comportant plus d'une couche cachée sont appelés réseaux neuronaux profonds. Les neurones de la couche cachée reçoivent des entrées de la couche d'entrée et fournissent des sorties à la couche de sortie.

Output Layer : La couche de sortie contient le nombre de neurones en fonction du nombre de classes de sortie. S'il s'agit d'un problème de classification multi-classes,

elle contient un nombre de neurones égal au nombre de classes. Pour une classification binaire, elle contient deux neurones.

Ce principe est repris pour les neurones formels. Un neurone artificiel se comporte comme une fonction mathématique réelle avec plusieurs variables et des valeurs réelles. Il possède donc plusieurs entrées et une sortie qui correspondent respectivement aux dendrites et à l'axone. Le modèle le plus simple appelé perceptron a été inventé en 1957 par F. Rosenblatt (voir figure 2.5) et est défini par différents éléments.

Un neurone j est défini par :

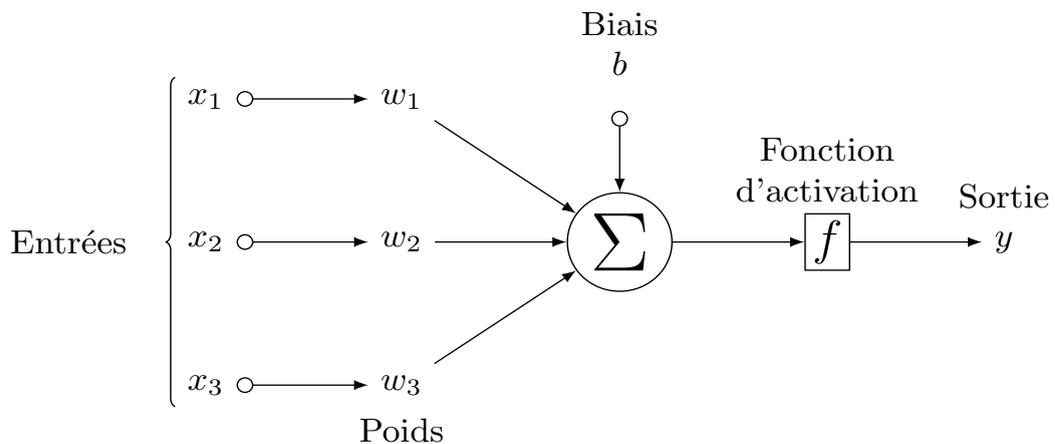


FIGURE 2.5 – Architecture d'un neurone artificiel (perceptron)

- des entrées : x_1, x_2, \dots, x_n
- des poids : w_1, w_2, \dots, w_n qui pondèrent les entrées.
- un biais : b_j qui régule l'activation du neurone.
- une fonction d'activation $\Phi()$: détaillée ci-dessous.
- une sortie : y_j

2.2.3 Les fonctions d'activation

Les fonctions d'activation sont un élément crucial du DL. Elles déterminent si un neurone se déclenche ou non. Pour apprendre une fonction complexe, la fonction d'activation doit être non linéaire [Xiao, 2019]. Dans ce qui suit, nous allons décrire

les différentes fonctions d'activation.

Dans la figure 2.5 , les entrées sont multipliées par des poids, puis transmises à la couche cachée suivante. Le biais est également donné en entrée avec les entrées pondérées. La somme pondérée passe par une fonction non linéaire appelée fonction d'activation.

2.2.3.1 Fonction d'activation

Les fonctions d'activation sont un élément crucial de l'apprentissage profond. Elle détermine si un neurone se déclenche ou non. Afin d'apprendre une fonction complexe, la fonction d'activation doit être non linéaire. Dans ce qui suit, nous allons décrire les différentes fonctions d'activation illustrées dans la figure 2.6.

- **Fonction Sigmoid (sigmoïde)** : L'objectif principal de cette fonction est de réduire la valeur d'entrée entre 0 et 1. Elle est définie comme suit :

$$F(x) = \frac{1}{1 + e^{-x}} \quad (2.1)$$

- **Fonction Tanh** : Également appelée "tangente hyperbolique", elle produit un résultat compris entre -1 et 1. Elle est définie comme suit :

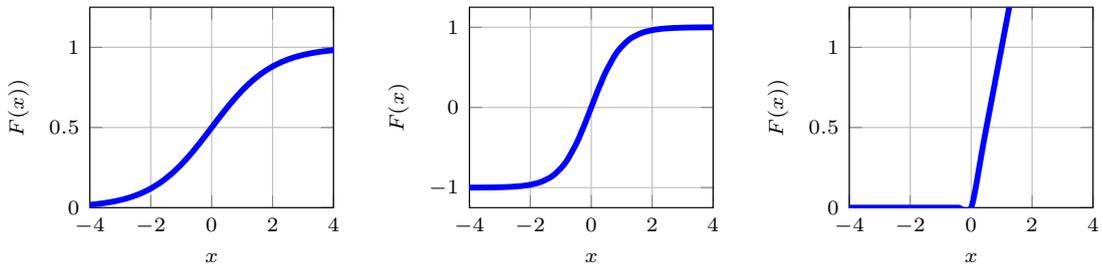
$$F(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2.2)$$

- **Unité linéaire rectifiée (ReLU)** : ReLU est la fonction la plus couramment utilisée en apprentissage profond augmente les propriétés de la fonction de décision et de l'ensemble du réseau sans affecter les champs réceptifs de la couche convolutive. Elle est définie comme suit :

$$F(x) = \max(0, x) \quad (2.3)$$

- **Fonction Softmax** : Dans les modèles d'apprentissage profond, la fonction SoftMax est la dernière couche qui est utilisée pour calculer les scores de la classe. L'entrée du classificateur SoftMax est un vecteur de caractéristiques résultant du processus d'apprentissage et la sortie est une probabilité qu'une image appartienne à une classe donnée . La fonction d'activation Softmax a la formule suivante.

$$F(x) = \frac{e^{x_i}}{\sum_{\geq 0} e^{x_p}} \quad (2.4)$$



(a) Sigmoid activation function (b) Hyperbolic tangent activation (c) Rectified linear unit activation

FIGURE 2.6 – Les fonctions d’activations

2.2.3.2 Convolutional Neural Network :

Un Convolutional Neural Network (CNN) ou bien un réseau neuronal convolutif est un type de réseau neuronal artificiel qui utilise des perceptrons multicouches, dans ce cas les couches cachées (Hidden Layers) peuvent être des couches convolutionnelles utilisées dans la reconnaissance et le traitement d’images [LeCun et al., 1989]. La figure 2.7 présente un exemple de CNN pour la classification des images (bouteille par exemple).

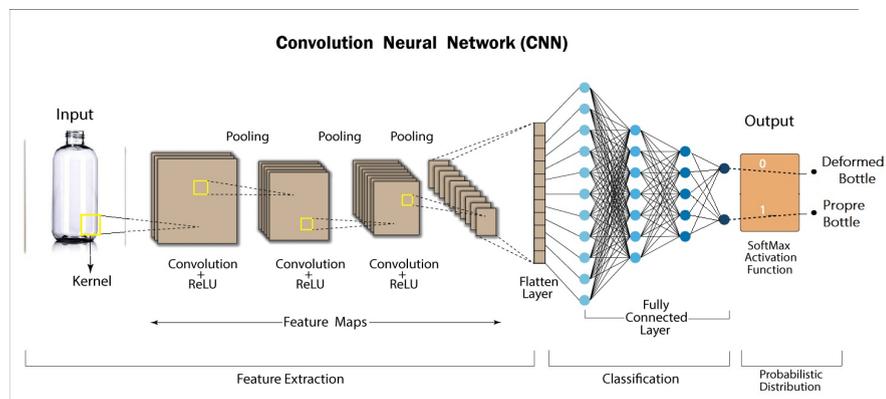


FIGURE 2.7 – Exemple du réseau de neurones convolutifs CNN 2D(classification d’une bouteille)

Un CNN est formé de :

- **Couche de convolution** : La couche convolutionnelle calcule l'opération de convolution des images d'entrée en utilisant des filtres à noyau pour extraire les caractéristiques fondamentales. Les filtres à noyau dans une couche de convolution sont de la même dimension mais avec des tailles constants plus petites que les images d'entrée. Le masque de filtrage K glisse sur l'ensemble de l'image d'entrée I étape par étape et estime le produit scalaire entre les poids des filtres à noyau avec la valeur de l'image d'entrée, ce qui permet de produire une carte F 2D donnée par l'équation :

$$F(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n)K(i - m, j - n) \quad (2.5)$$

Ainsi, le CNN apprend une caractéristique visuelle. La figure 2.8 illustre un exemple simple du calcul dans un CNN.

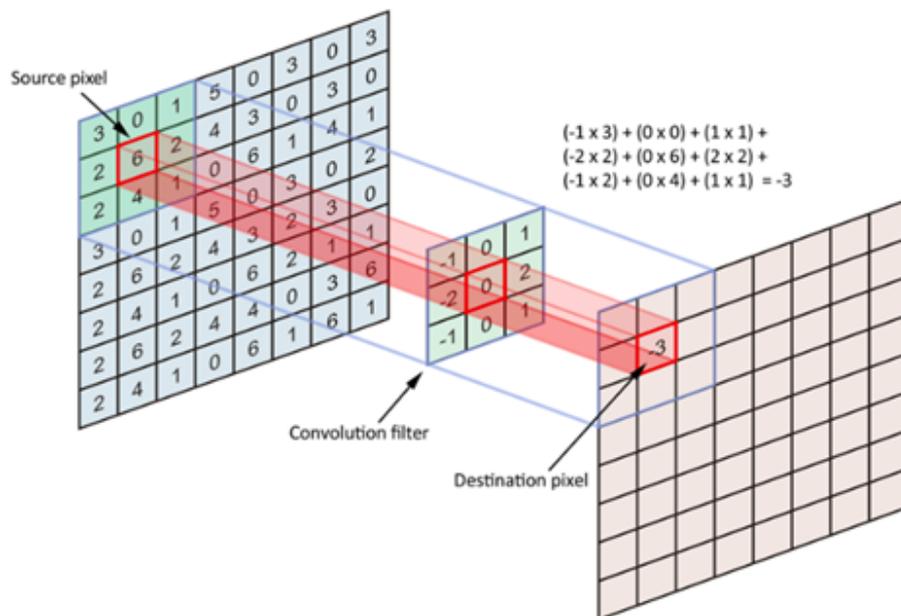


FIGURE 2.8 – Couche de convolution 2D

- **Pooling layer (couche de pooling)** : La couche qui vient juste après chaque couche de convolution est appelée couche de pooling. Une couche de pooling permet de réduire la dimension de la carte d'activation. Elle constitue normalement une entrée pour la couche de convolution suivante obtenue par le processus de sous-échantillonnage. Elle conduit à la perte de certaines informations, ce qui aide le réseau à réduire à la fois l'overfitting et la charge de

calcul. Une fenêtre choisie va glisser sur toute la zone de l'entrée, étape par étape, en la convertissant en une valeur spécifique basée sur une méthode de sous-échantillonnage. L'approche de sous-échantillonnage fonctionne en utilisant une approche de moyenne ou de valeur maximale [Lee et al., 2016].

En raison de ses excellentes performances, nous préférons choisir l'approche de la valeur maximale. La figure 2.9 représente une opération simple de réduction de la dimension d'une carte d'activation à l'aide d'une approche de sous-échantillonnage maximal.

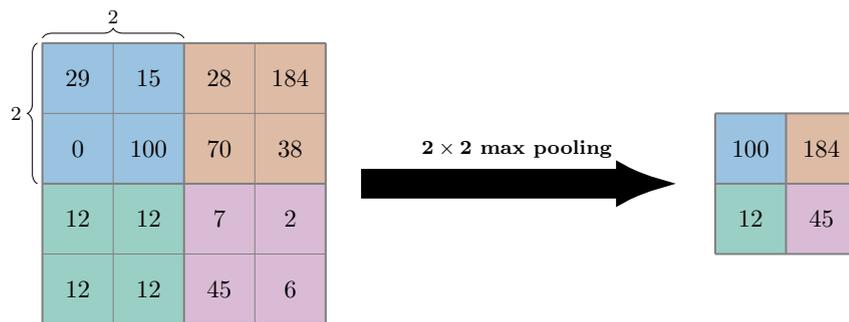


FIGURE 2.9 – Max Pooling

- **Fully connected layer (couche entièrement connectée)** : La couche entièrement connectée se trouve à la fin du réseau neuronal et est utilisée à des fins de classification. Elle prend un ensemble de données d'entrée des couches précédentes et analyse globalement la sortie de toutes les couches précédentes. Contrairement à la couche de convolution, il s'agit d'une opération globale qui utilise des fonctions d'activation telles que l'activation SoftMax et la fonction d'activation Sigmoid pour classifier les images d'entrée. La plupart du temps, la fonction d'activation SoftMax est utilisée pour la classification. La figure 2.10 donne un exemple d'une couche entièrement connectée.
- **Dropout** : Habituellement, lorsque toutes les caractéristiques sont connectées à la couche Fully connected, cela peut entraîner un overfitting (sur-ajustement) dans l'ensemble de données d'apprentissage. Le overfitting se produit lorsqu'un modèle particulier fonctionne si bien sur les données d'apprentissage qu'il a un impact négatif sur les performances du modèle lorsqu'il est utilisé sur de nouvelles données.

Pour surmonter ce problème, on utilise une couche d'exclusion dans laquelle quelques neurones sont éliminés (voir la figure 2.11) du réseau neuronal pen-

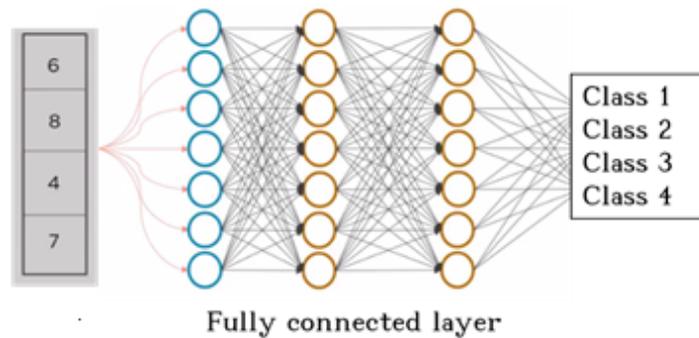


FIGURE 2.10 – Exemple d’une couche entièrement connectée

dant le processus d’apprentissage, ce qui réduit la taille du modèle.

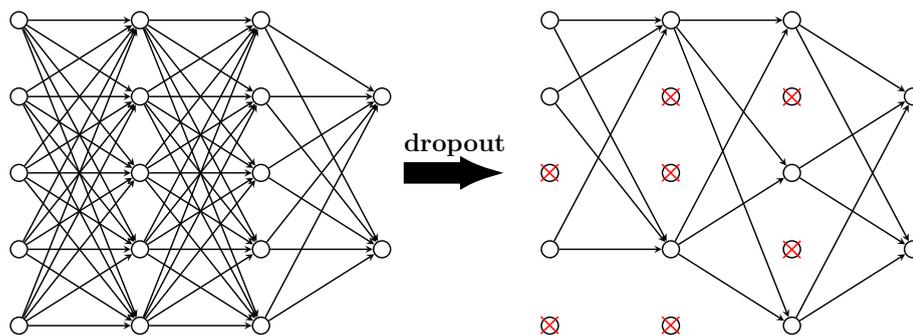


FIGURE 2.11 – la différence entre original networks et dropout networks

- **Flattening Layer (couche d’aplatissement)** : Les données d’entrée de la dernière couche, la couche entièrement connectée, sont préparées dans cette couche. Les réseaux neuronaux artificiels utilisent généralement un tableau unidimensionnel comme données d’entrée. . Dans cette couche, les données de la couche précédente sont converties en un tableau unidimensionnel comme l’indique la figure 2.12.

2.2.4 Apprentissage d’un réseau de neurones convolutifs (Neuron learning) :

Le processus d’ajustement de la valeur des poids est défini comme le "training " ou apprentissage du réseau neuronal. Premièrement, le CNN démarre avec des poids aléatoires. Pendant la l’apprentissage du CNN, le réseau neuronal est alimenté par

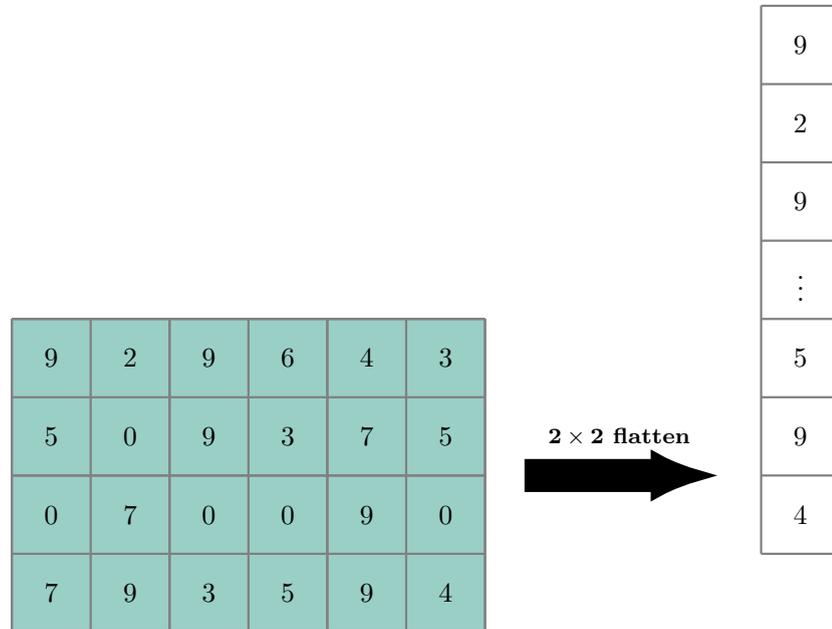


FIGURE 2.12 – Exemple de flattening 2D

un grand ensemble de données d’images étiquetées avec leurs étiquettes de classe correspondantes (chat, chien, cheval, automobiles, bouteilles, etc.). Le réseau CNN traite chaque image en lui attribuant des valeurs de manière aléatoire, puis effectue des comparaisons avec l’étiquette de classe de l’image d’entrée. Si la sortie ne correspond pas à l’étiquette de classe le réseau CNN procède à un petit ajustement des poids de ses neurones CNN afin que la sortie corresponde correctement à l’étiquette de classe de l’image.

Le processus d’apprentissage s’effectue en deux étapes : Forward et Backward.

L’étape Forward consiste à déterminer une prédiction faite par le neurone et à calculer l’erreur par rapport à une valeur de référence indiquée par les données d’apprentissage. L’étape Backward met à jour les poids des neurones en fonction de l’erreur effectuée par l’algorithme d’optimisation choisi .

Différentes fonctions de coût sont définies telles que :

- **Erreur quadratique moyenne :**

$$L = \frac{1}{n} \left(\sum_{i=1}^n (y^i - Y^i)^2 \right) \quad (2.6)$$

- **Entropie croisée :**

$$L = \frac{1}{n} \sum_{i=1}^n [y^i \log(Y^i) + (1 - y^i) \log(1 - Y^i)] \quad (2.7)$$

où Y est la prédiction du neurone, y est la valeur de référence des données d'apprentissage et n est le nombre de prédictions.

La descente de gradient est un algorithme d'optimisation utilisé pour minimiser (ou maximiser) une fonction objectif. En Machine Learning, la fonction objectif correspond à l'erreur d'apprentissage de dimension d où d est le nombre de paramètres du modèle.

L'utilisation de la descente du gradient n'est pas réalisable lorsque la quantité de données est importante.

Pour surmonter ce problème, l'ensemble des données est divisé en mini-batches. Pour chaque mini-batch, la perte et le gradient sont calculés et les paramètres sont mis à jour.

Cette stratégie appelée Stochastic Gradient Descent (SGD) est la plus utilisée aujourd'hui en raison de ses performances. Elle offre un bon compromis entre la robustesse et l'évolution rapide du réseau. En plus de l'algorithme SGD, il existe plusieurs autres algorithmes utilisés pour mettre à jour les paramètres et ainsi minimiser la perte tels que l'algorithme ADAM et l'algorithme RMSprop.

Un réseau neuronal peut être entraîné des dizaines, des centaines, voire des milliers d'itérations pour améliorer sa précision. Chaque passage sur l'ensemble des données pour entraîner le réseau neuronal est appelé une époque.

Un problème majeur inclus dans le processus d'apprentissage est la sélection de l'algorithme d'apprentissage (optimiseur). L'optimiseur Adam est une technique d'optimisation qui remplace la procédure classique de descente de gradient stochastique. L'optimisation suit une méthode itérative pour mettre à jour les nouveaux poids pendant l'apprentissage. Le processus est simple à suivre et efficace en termes de calcul. Il consomme moins de mémoire et a également une bonne performance.

Le nombre d'époques (epoch) est le nombre de fois que l'ensemble du processus d'apprentissage passe par le réseau neuronal. Une époque signifie que chaque échantillon de l'ensemble de données de d'apprentissage a eu l'occasion de mettre à jour les paramètres du modèle interne.

La taille du lot (batch size) est un hyperparamètre qui définit le nombre d'échan-

tillons à traiter avant de mettre à jour les paramètres du modèle interne.

Le réglage des hyperparamètres consiste à choisir un ensemble d'hyperparamètres adapté et optimal pour l'algorithme d'apprentissage. Les hyperparamètres sont des paramètres dont la valeur est utilisée pour contrôler le processus d'apprentissage et est définie avant le début du processus d'apprentissage. En général, les valeurs des paramètres de poids des nœuds sont en cours d'apprentissage. Ce réglage se fait selon deux stratégies : La recherche par grille et la recherche aléatoire.

2.3 Conclusion

Dans ce chapitre, nous avons présenté les types de l'apprentissage automatique. Le concept du deep learning a été introduit par la suite en mettant l'accent sur les notions de bases liées à ce sujet et qui seront utilisées dans le prochain chapitre.

Chapitre 3

Classification Automatique des Bouteilles en plastique pour le Contrôle de la Qualité

Sommaire

3.1	Introduction	35
3.2	Travaux connexes	35
3.3	Le contrôle de la qualité des bouteilles en utilisant le CNN	38
3.3.1	Base de données et protocole d'évaluation	39
3.4	Conclusion	42

3.1 Introduction

Les fabricants d'aliments et de boissons sont tenus de contrôler la qualité de leurs produits (questions réglementaires). La mise en circulation de produits (et/ou d'emballages) défectueux peut donner lieu à des plaintes, voire le retour de toute une série de produits (notamment dans les industries alimentaire et pharmaceutique). Pour réduire le risque de perte financière et de réputation, les fabricants installent des systèmes pour contrôler tous leurs produits [Krones, 2016]. Une approche largement utilisée pour le contrôle de la qualité consiste à appliquer des systèmes basés sur la vision artificielle et le traitement des images [Czimmermann et al., 2020, Golnabi and Asadpour, 2007, Luster and Batchelor, 2012].

Les avantages fondamentaux de ces solutions, tels que l'efficacité, l'absence de contact, ou un degré de complexité relativement faible, en font un choix fréquent pour le contrôle de la qualité dans l'industrie alimentaire [Brosnan and Sun, 2004, Patel et al., 2012]. Les systèmes basés sur l'image ne peuvent être utilisés que lorsqu'il est possible de distinguer les produits défectueux des produits acceptables sur la base d'images (dans une certaine gamme d'ondes électromagnétiques : ultraviolet, lumière visible, infrarouge et rayons X).

3.2 Travaux connexes

Au cours des dernières années, diverses activités de recherche ont été menées afin de proposer des systèmes de vision artificielle intelligents pour l'inspection des produits défectueux, basés sur l'exploitation des données générées par différentes technologies intégrées dans les lignes de fabrication modernes, en utilisant diverses techniques d'apprentissage automatique. Voici quelques réalisations exceptionnelles dans ce domaine.

[Sahoo et al., 2019] ont proposé une structure d'inspection dynamique des bouteilles. Tout d'abord, l'image de la bouteille à inspecter a été capturée par une caméra intelligente à haute résolution. Ensuite, des techniques de traitement d'image ont été utilisées pour réduire le bruit et améliorer la qualité des images capturées. La région d'intérêt (parties défectueuses de la bouteille) a été sélectionnée dans l'image originale par recadrage, puis convertie en une image en niveaux de gris. L'image

de la bouteille a été segmentée du fond de l'image en appliquant des méthodes de segmentation. Ensuite, un ensemble de données a été construit à partir des caractéristiques adaptatives estimées en utilisant des concepts mathématiques tels que le vecteur caractéristique bidimensionnel moyen en niveaux de gris, la transformée en ondelettes et l'Analyse en Composantes Principales (ACP). Un réseau neuronal artificiel (ANN) entraîné par l'algorithme de rétro propagation (BP), l'algorithme d'évaluation différentielle (DEA) et une machine à vecteur de support (SVM) a été utilisé pour classer les images en considérant les caractéristiques extraites comme variables d'entrée et les types de bouteilles défectueuses comme variables de sortie. Au cours de l'inspection, il existe un risque d'éclairage insuffisant et de changement de position d'un objet en raison de divers facteurs, ce qui signifie qu'un dispositif sensoriel doit être mis en œuvre. Les auteurs de cet article ont comparé les différents algorithmes de classification utilisant trois méthodes d'extraction de caractéristiques (vecteur de caractéristiques 2D en niveaux de gris AVG, caractéristiques basées sur l'ACP et caractéristiques basées sur les ondelettes) avec et sans mise en œuvre d'un capteur dans le système d'inspection par vision industrielle. Les résultats obtenus ont révélé que le temps de calcul de l'ANN utilisant la DEA est comparativement inférieur à celui de l'ANN utilisant la BP. Le temps de calcul moyen du système proposé sans et avec l'utilisation de capteurs est de 33,04 et 29,80, respectivement. La comparaison du temps de calcul moyen indique que le temps de calcul est plus important pour la classification des bouteilles sans capteur que pour la classification avec capteur. Le système suggéré est capable de classer les imperfections des bouteilles avec un taux de succès de classification de 91,25 à 97,5%. Finalement, il a été conclu que le système d'inspection par vision artificielle basé sur l'IA est le plus approprié pour l'inspection du niveau de qualité des imperfections des bouteilles.

Wang et al. [Wang et al., 2019] ont proposé un modèle de vision artificielle pour l'inspection des produits défectueux basé sur l'apprentissage profond et la transformée de Hough. La méthode conçue détecte les produits défectueux à travers trois étapes principales : le pré-traitement de l'image, l'extraction de la région d'intérêt (ROI) et l'identification de l'image. Dans un premier temps, un filtre gaussien a été effectué sur l'image acquise pour limiter le bruit aléatoire et obtenir une meilleure représentation de l'image brute. Le fonctionnement continu du dispositif d'imagerie conduit inévitablement à une augmentation excessive de sa température. Par conséquent, les images acquises seront corrompues par le bruit gaussien. Par conséquent,

un filtre gaussien a été utilisé comme étape de pré-traitement afin d'améliorer les structures de l'image et de réduire le bruit. Ensuite, la transformée de Hough probabiliste a été utilisée pour extraire le ROI des images filtrées. Les images capturées ne garantissent pas que toutes les informations contenues soient liées à la tâche. Par exemple, le contenu de l'arrière-plan peut occuper une partie considérable de l'image, ce qui entraîne des calculs inutiles lors de l'étape d'identification suivante. Par conséquent, la technique de la transformée de Hough a été utilisée pour isoler les caractéristiques et extraire la ROI des images filtrées. Enfin, les auteurs ont utilisé un réseau neuronal convolutif pour construire le module d'identification. Ils ont construit le module d'identification d'image sur la base du bloc résiduel inversé, ce qui constitue une bonne alternative pour réduire la taille du modèle et le temps de calcul. L'identification d'image a été réalisée en utilisant la méthode proposée avec un grand nombre de jeux de données, qui consistaient en des images de bouteilles défectueuses et non défectueuses. L'analyse des résultats obtenus a révélé que la préparation des données est une étape cruciale et décisive pour le succès de tout modèle. L'application de la méthode d'extraction du retour sur investissement basée sur la transformée de Hough probabiliste avec l'aide d'une source lumineuse rectangulaire a permis d'obtenir la région correcte de la bouteille dans l'image originale. Par conséquent, le temps nécessaire à l'identification de l'image a été considérablement réduit. La méthode d'inspection considérée passe en moyenne 47,60 ms sur chaque identification d'image et atteint une précision de 99,60%. L'évaluation des résultats a confirmé que le modèle proposé dans cet article est capable de satisfaire les besoins de la plupart des lignes de production, en considérant que le système a le potentiel d'obtenir des résultats plus précis au fur et à mesure que des échantillons défectueux sont découverts pendant la production et utilisés pour l'entraînement du réseau.

Travaux antérieurs sur le système d'inspection des bouchons de bouteilles Nous avons effectué des recherches sur les études antérieures concernant le système d'inspection des bouchons de bouteilles. Leurs différentes méthodes pour entreprendre cette tâche consistent en un traitement d'image et vision artificielle. Une étude menée par Ren et al, [Ren et al., 2020] montre un excellent résultat en utilisant l'approche basée sur le traitement d'image. Ils ont corrigé le flou des images causé par le fonctionnement à grande vitesse des bandes transporteuses en appliquant une technique d'amélioration du flou et d'accentuation des bords basés sur la méthode Prewitt, ce

qui permet d'améliorer la précision. Le système met également l'accent sur la détection des bouchons dans une production à grande échelle, l'automatisation traite les images à une à une vitesse de 120/min. Une autre étude réalisée par Pruthvi Kumar et Ramakrishna [Kumar and Ramakrishna, 2015] montre que l'utilisation de la vision industrielle en remplacement un système d'inspection conventionnel basé sur des capteurs améliore la qualité de l'inspection, l'utilisation de l'image et la productivité, utilisable avec une variété de produits. Ils ont réalisé l'étude en utilisant détection des bords et des méthodes de regroupement d'images pour segmenter et identifier les parties du bouchon de la bouteille. Un autre article sur un système basé sur la vision. [Zhou et al., 2014] proposent une technique de combinaison de la représentation éparsée et de l'histogramme de projection de région circulaire (CRPH) pour créer un système de détection rapide des défauts sur la surface des capsules de bouteilles (bouchons de bouteilles).

3.3 Le contrôle de la qualité des bouteilles en utilisant le CNN

La figure 3.1 montre une vue d'ensemble du réseau de neurones convolutifs que nous utilisons. Il peut être divisé en 11 couches décrites comme suite :

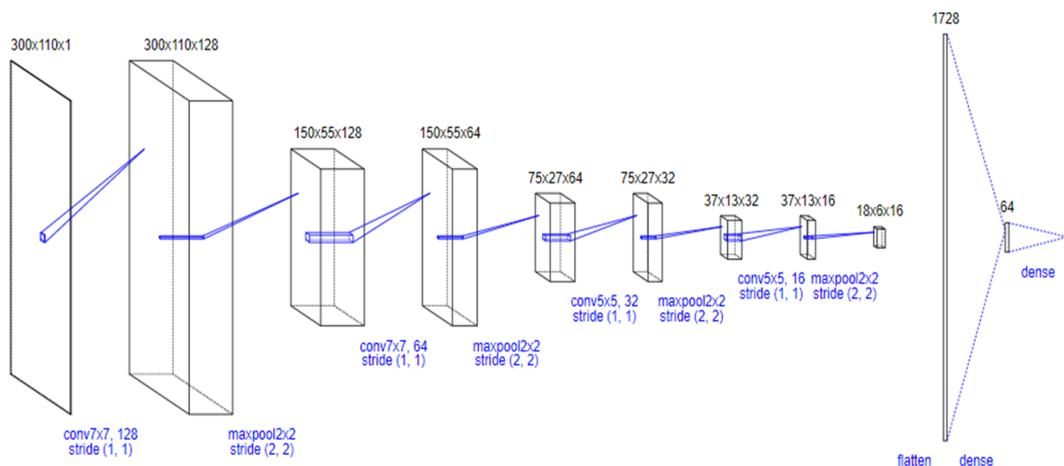


FIGURE 3.1 – Architecture du modèle de contrôle de qualité

- La couche 1 est constituée d'une couche convolutive avec fonction d'activation ReLu. Cette couche couche reçoit une image de taille 300×110 . La taille du

filtre convolutif est de 7×7 , le padding est same est le stride est de 1 et le nombre de filtre est de 128.

- La couche 2 est la couche max pooling. Cette couche reçoit une entrée de taille $128@224 \times 224$ de la couche précédente. La taille du pooling= 2×2 , le padding = 0 et le stride= 2 Cette couche n'a pas de fonction d'activation.
- La couche 3 est constituée d'une couche convolutive avec fonction d'activation ReLu. Cette couche couche reçoit une entrée de taille $128@150 \times 55$. La taille du filtre convolutif est de 7×7 , le padding est same est le stride est de 1 et le nombre de filtre est de 64.
- La couche 4 est la couche max pooling. Cette couche reçoit une entrée de taille $64@150 \times 50$ de la couche précédente. La taille du pooling= 2×2 , le padding = 0 et le stride= 2.
- Le vecteur de caractéristiques issu des convolutions a une dimension de 1728 obtenu par la couche flatten.
- L'avant dernière couche du réseau est une couche entièrement connectée composée de 64 noeuds avec fonction d'activation ReLu.
- La dernière couche du réseau est une couche entièrement connectée composée de 2 noeuds avec fonction d'activation Softmax. Cette couche calcule les scores des classes, résultant en un vecteur de taille 2, où chacun correspond à un score de classe.

3.3.1 Base de données et protocole d'évaluation

Le développement du modèle de contrôle de la qualité des bouteilles commence par la collecte des données. Plus précisément, nous avons utilisé d'un ensemble de données de 1000 images de bouteilles d'eau normales et défectueuses.

La figure 3.2 montre quelques exemples de cette base.

La base des données est divisée selon un ratio 80 : 20, ce qui signifie que 80% des données sont prises pour l'apprentissage et 20% pour le test (voir le listing 3.1).

```
1 from sklearn.model_selection import train_test_split
2 X_train,X_test,y_train,y_test= train_test_split(X,y,test_size =
   0.2,random_state = 42)
3 print(X_train.shape)
4 print(y_train.shape)
5 print(X_test.shape)
```



FIGURE 3.2 – Exemple de bouteilles défectueuses et non défectueuses

```
6 print(y_test.shape)
```

Listing 3.1 – Répartition de la base de données

Une fois le modèle créé et la base de données répartie, des paramètres d'initialisation tels que la taille du batch, le nombre d'époques, le taux d'apprentissage, doivent être ajustés pour obtenir de meilleures performances. La taille du batch est le nombre d'images dans un lot (batch). Le nombre d'époques se définit comme étant le nombre de fois que toutes les données d'apprentissage sont passées par le réseau. Le taux d'apprentissage est un paramètre très sensible qui pousse le modèle vers la convergence. La recherche de sa meilleure valeur se fait par un processus expérimental. Dans ce travail, l'étape d'apprentissage est effectuée avec des hyperparamètres définis comme `Batch_size = 32` et `Epochs = 20` (voir le listing 3.2). Enfin, le modèle est compilé en appelant la fonction `compile()`, qui a comme entrées :

- **optimiser** utilisé pour optimiser le réseau neuronal, tel que Adagrad et Adam.
- **loss** qui fait référence à la valeur que le modèle doit minimiser, comme l'entropie croisée et l'erreur quadratique moyenne.
- **Metrics** qui évaluent la précision des problèmes de classification.

```
1 model.compile(loss='categorical_crossentropy',
```

```

2         optimizer='adam',
3         metrics=['accuracy'])
4 history=model.fit(X_train,y_train,batch_size=32,epochs=20,verbose
                    =1)

```

Listing 3.2 – Hyper paramètres et compilation du modèle

Le modèle est entraîné à l'aide de la méthode `fit()`. Cette méthode permet au modèle d'itérer sur les données et de trouver le réseau neuronal le plus optimal pour ces données. Une matrice de confusion est un outil qui permet de visualiser les performances d'un algorithme utilisé dans l'apprentissage supervisé. Dans un problème à deux classes, le modèle de classification prédit l'une des deux classes qui sont généralement désignées sous le nom de classes positive et négative. Étant donné un classificateur et un exemple à classer, quatre situations sont envisageables. Si l'exemple est positif et qu'il est classé comme positif, il compte comme un vrai positif TP ; s'il est classé négatif, il compte comme un faux négatif FN . Si l'exemple est négatif et qu'il est classé comme négatif, il compte comme un vrai négatif TN ; s'il est classé positif, il compte comme un faux positif FP . Les sorties produites par le classificateur peuvent être représentées par une matrice de confusion de 2×2 . La Table 3.1 montre la matrice de confusion pour un problème à deux classes où les lignes indiquent la classe réelle et les colonnes indiquent la classe prédite.

TABLE 3.1 – Matrice de confusion d'un problème à deux classes

		Classe prédite	
		Positive	Negative
Classe réelle	Positive	TP	FN
	Negative	FP	TN

Le taux de bonne classification est la mesure de performance la plus simple et qui est à la base de tout autre critère. Le taux de classification calcule le pourcentage d'échantillons correctement classés par rapport au nombre total d'échantillons. À partir des valeurs de la matrice de confusion (voir la Table 3.1), le taux de classification est donné par l'équation suivante :

$$TC = \frac{TP + TN}{TP + FN + TN + FP} \quad (3.1)$$

3.3.1.1 Résultats obtenus

Parmi les images de test pour la classification de la qualité de la bouteille, notre modèle atteint un taux de classification de 97.35%. Les performances du classifieur sont analysées à l'aide de la matrice illustré dans la figure 3.3.

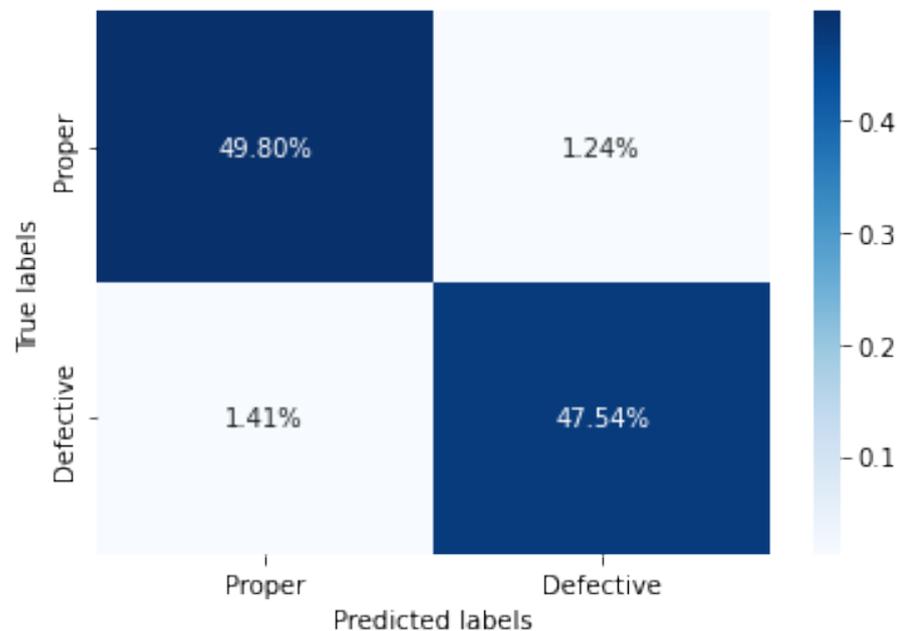


FIGURE 3.3 – La matrice de confusion de la qualité

On remarque que le modèle basé sur le CNN a très peu d'erreurs de classification. Ce modèle classe les bouteilles défectueuses comme non défectueuses avec un taux de 1.41% et classe les bouteille non défectueuse comme défectueuses avec un taux de 1.24%.

La figure 3.4 montre l'évolution du taux d'apprentissage et de l'erreur en fonction du nombre d'époques.

Quelques résultats sont présentés dans la figure 3.5

3.4 Conclusion

Dans ce chapitre, nous avons présenté le modèle basé sur les réseaux de neurones convolutifs. Des résultats satisfaisants ont été obtenus prouvant le robustesse des CNN pour le contrôle intelligent de la qualité.

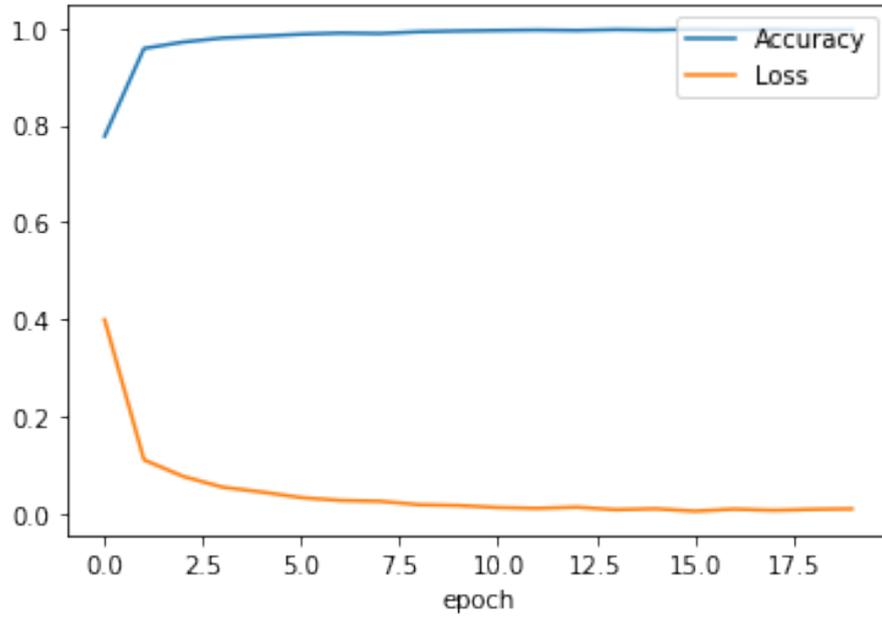


FIGURE 3.4 – Le taux d’apprentissage e l’erreur en fonction du nombre d’époques

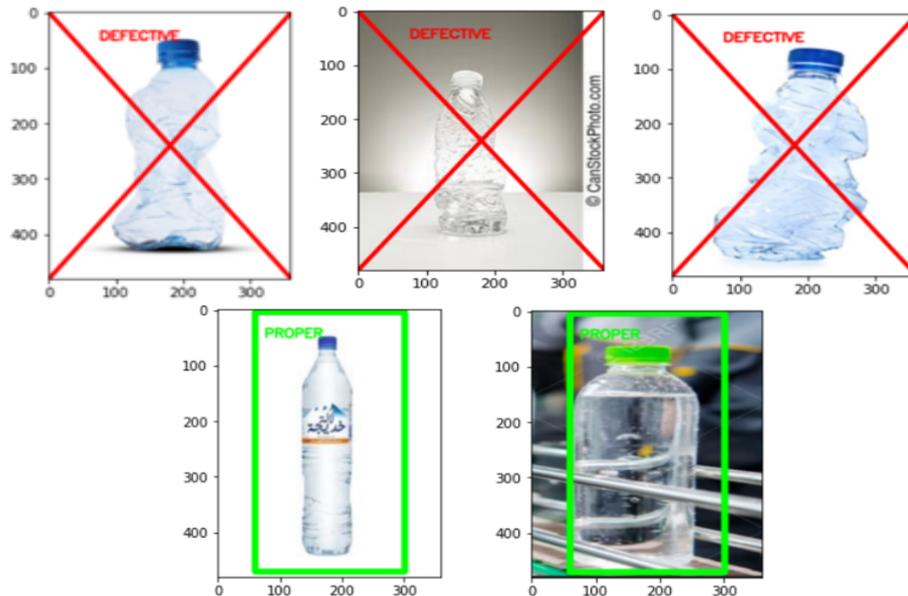


FIGURE 3.5 – Exemple de bouteilles défectueuses et non défectueuses

Conclusion Générale

Les systèmes de vision basé sur un ordinateur permettent de réaliser un grand nombre d'applications dans le domaine de l'automatisation à faible coût, en particulier dans les processus d'inspection où il est nécessaire d'améliorer la qualité de l'inspection. La programmation de ce système développé est flexible, rapide. Cela augmentera le contrôle de la qualité totale et éliminera la main d'œuvre des inspecteurs et dans certaines applications, ils doivent parfois évoluer dans des environnements spéciaux, dangereux et non compatibles avec les humains. Il permet d'augmenter la productivité, d'améliorer la qualité de l'inspection, la variété des produits et la rentabilité.

Le travail qui vient d'être réalisé propose une inspection automatisée de la qualité d'un produit à l'aide d'un système de vision artificielle et apprentissage profond qui est le réseau neuronal à convolution (CNN), que nous avons entraînés pour vérifier la qualité de la bouteille

Bibliographie

- [FPS,] What does fps mean? <https://clipchamp.com/en/definition/what-does-fps-mean-frame-rates-explained/>.
- [Andrews et al., 2013] Andrews, H. C., Billingsley, F., Fiasconaro, J., Frieden, B., Read, R., Shanks, J., and Treitel, S. (2013). *Picture processing and digital filtering*, volume 6. Springer Science & Business Media.
- [Bishop and Nasrabadi, 2006] Bishop, C. M. and Nasrabadi, N. M. (2006). *Pattern recognition and machine learning*, volume 4. Springer.
- [Brosnan and Sun, 2004] Brosnan, T. and Sun, D.-W. (2004). Improving quality inspection of food products by computer vision—a review. *Journal of food engineering*, 61(1) :3–16.
- [Castleman, 1996] Castleman, K. R. (1996). *Digital image processing*. Prentice Hall Press.
- [Czimmermann et al., 2020] Czimmermann, T., Ciuti, G., Milazzo, M., Chiurazzi, M., Roccella, S., Oddo, C. M., and Dario, P. (2020). Visual-based defect detection and classification approaches for industrial applications—a survey. *Sensors*, 20(5) :1459.
- [Golnabi and Asadpour, 2007] Golnabi, H. and Asadpour, A. (2007). Design and application of industrial machine vision systems. *Robotics and Computer-Integrated Manufacturing*, 23(6) :630–637.
- [Krones, 2016] Krones, A. (2016). Collaborative research effort between industry week custom research and kronos incorporated : The future of manufacturing : 2020 and beyond. In *Technical Report*. Krones.

- [Kumar and Ramakrishna, 2015] Kumar, S. and Ramakrishna, H. (2015). Automated bottle cap inspection using machine vision system. *International Journal of Innovative Research in Technology*, 2(2) :131–136.
- [LeCun et al., 1989] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4) :541–551.
- [Lee et al., 2016] Lee, C.-Y., Gallagher, P. W., and Tu, Z. (2016). Generalizing pooling functions in convolutional neural networks : Mixed, gated, and tree. In *Artificial intelligence and statistics*, pages 464–472. PMLR.
- [Luster and Batchelor, 2012] Luster, S. D. and Batchelor, B. G. (2012). Telecentric, fresnel and micro lenses. In *Machine vision handbook*, pages 259–281. Springer.
- [Messaoudi, 2020] Messaoudi, Z. (2020). Les 3 étapes essentielles de l’apprentissage automatique (machine learning). <https://www.spiria.com/fr/blogue/intelligence-artificielle/3-etapes-essentielles-apprentissage-automatique-machine-learning/>.
- [Mitchell and Mitchell, 1997] Mitchell, T. M. and Mitchell, T. M. (1997). *Machine learning*, volume 1. McGraw-hill New York.
- [Patel et al., 2012] Patel, K. K., Kar, A., Jha, S., and Khan, M. (2012). Machine vision system : a tool for quality inspection of food and agricultural products. *Journal of food science and technology*, 49(2) :123–141.
- [Petrou and Petrou, 2010] Petrou, M. M. and Petrou, C. (2010). *Image processing : the fundamentals*. John Wiley & Sons.
- [Ren et al., 2020] Ren, X., Wen, J., Lan, Y., Li, T., and Wang, X. (2020). Design of bottle cap detection system based on image processing. In *2020 Chinese Control And Decision Conference (CCDC)*, pages 4880–4885. IEEE.
- [Sahoo et al., 2019] Sahoo, S. K., Mahesh Sharma, M., and Choudhury, B. (2019). A dynamic bottle inspection structure. In *Computational Intelligence in Data Mining*, pages 873–884. Springer.
- [Thévenaz et al., 2000] Thévenaz, P., Blu, T., and Unser, M. (2000). Image interpolation and resampling. *Handbook of medical imaging, processing and analysis*, 1(1) :393–420.

- [Wang et al., 2019] Wang, J., Fu, P., and Gao, R. X. (2019). Machine vision intelligence for product defect inspection based on deep learning and hough transform. *Journal of Manufacturing Systems*, 51 :52–60.
- [Xiao, 2019] Xiao, Y. (2019). Vehicle detection in deep learning. *arXiv preprint arXiv :1905.13390*.
- [Zhou et al., 2014] Zhou, W., Fei, M., Zhou, H., and Li, K. (2014). A sparse representation based fast detection method for surface defect detection of bottle caps. *Neurocomputing*, 123 :406–414.