



INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO
AMAZONAS
CAMPUS MANAUS-DISTRITO INDUSTRIAL
CURSO DE BACHARELADO EM ENGENHARIA DE CONTROLE E
AUTOMAÇÃO

VITOR ARLINSON RODRIGUES DA SILVA

**MÉTODOS PARA RECONHECIMENTO DE RESÍDUOS RECICLÁVEIS
ATRAVÉS DE VISÃO COMPUTACIONAL**

MANAUS - AM

2025

VITOR ARLINSON RODRIGUES DA SILVA

**MÉTODOS PARA RECONHECIMENTO DE RESÍDUOS RECICLÁVEIS
ATRAVÉS DE VISÃO COMPUTACIONAL**

Trabalho de Conclusão de Curso apresentado ao Instituto Federal de Educação, Ciência e Tecnologia do Amazonas, Campus Manaus Distrito-Industrial, Curso de Bacharelado em Engenharia de Controle e Automação para obtenção do título de Bacharel em Engenharia de Controle e Automação.

Orientador: Prof. Dr. Alyson de Jesus dos Santos

Coorientador: Profa. Esp. Micila Sumaria Medeiros Pereira

MANAUS - AM

2025

Dados Internacionais de Catalogação na Publicação (CIP)

S586m Silva, Vitor Arlinson Rodrigues da.
Métodos para reconhecimento de resíduos recicláveis através de visão computacional / Vitor Arlinson Rodrigues da Silva. — Manaus, 2025.
49f.: il. color.

Monografia (Graduação) — Instituto Federal de Educação, Ciência e Tecnologia do Amazonas, *Campus* Manaus Distrito Industrial, Curso de Engenharia de Controle e Automação, 2025.

Orientador: Prof.º Alyson de Jesus dos Santos, Dr.

Coorientadora: Prof.^a Micila Sumária Medeiros Pereira, Esp.

1. Visão computacional. 2. Classificador. 3. Resíduos recicláveis. I. Santos, Alyson de Jesus dos. II. Pereira, Micila Sumária Medeiros. III. Instituto Federal de Educação, Ciência e Tecnologia do Amazonas. IV. Título.

CDD 629.89

Elaborada por Oziane Romualdo de Souza (CRB11/ nº 734).

ANEXO 7

ATA DE DEFESA PÚBLICA DO TRABALHO DE CONCLUSÃO DE CURSO

Aos 10 dias do mês de Fevereiro, de 2025, às 14:30h, o(a) discente Vitor Arlison Rodrigues da Silva apresentou o seu Trabalho de Conclusão de Curso para avaliação da Banca Examinadora constituída pelos seguintes integrantes: Prof(a). Dr. Alyson de Jesus dos Santos (docente-orientador), Prof(a). Prof(a). Msc. Edevaldo Albuquerque Fialho (Membro 1) e Prof(a). Msc. Renan Cavalcante Santos (Membro 2). A sessão pública de defesa foi aberta pelo(a) presidente da banca, que apresentou a Banca Examinadora e deu continuidade aos trabalhos, fazendo uma breve referência ao TCC, que tem como título “Métodos para Detecção de Resíduos Recicláveis utilizando Visão Computacional”. Na sequência, o(a) discente teve até 30 minutos para a comunicação oral de seu trabalho. Cada integrante da banca examinadora fez suas arguições após a defesa do mesmo. Ouvidas as explicações do(a) discente, a banca examinadora, reunida em caráter sigiloso, para proceder à avaliação final, deliberou e decidiu pela APROVAÇÃO com média final 8,7 (Oito, Sete) do referido trabalho.

Foi dada ciência ao(à) discente que a versão final do trabalho deverá ser entregue até o dia 12/02/2025, com as devidas alterações sugeridas pela banca. Nada mais havendo a tratar, a sessão foi encerrada às 16h 30 min, sendo lavrada a presente ata, que, uma vez aprovada, foi assinada por todos os membros da Banca Examinadora e pelo(a) discente.

Prof. (a) Orientador(a)/Presidente: Alyson de Jesus dos Santos

Prof. (a) Avaliador 1: Edevaldo Albuquerque Fialho

Prof. (a) Avaliador 2: Renan Cavalcante Santos

Discente: Vitor Arlison Rodrigues da Silva

DEDICATÓRIA

Dedico este trabalho a minha família, minha mãe Eclene, meu pai Arlindo, meu irmão Dhiego e minha irmã Diane. Que sempre estiveram comigo, me acompanhando e dando todo o suporte para que durante o período acadêmico, pudesse focar nos estudos. Sem eles, nada eu seria.

AGRADECIMENTOS

O curso de Engenharia de Controle e Automação é um grande desafio, exigindo dedicação e resiliência. No entanto, o apoio e incentivo daqueles que estão ao nosso redor são fundamentais para alcançar o sucesso.

Gostaria de expressar minha profunda gratidão aos meus pais, que desde o início estiveram ao meu lado, oferecendo suporte incondicional, tanto emocional quanto financeiro. À minha irmã, Diane, que nunca mediu esforços para me proporcionar o melhor, e ao meu irmão, Dhiego, que, desde o princípio, esteve presente, aconselhando-me nos momentos mais difíceis.

Um agradecimento especial à minha avó Eclair. Embora ela não esteja aqui para celebrar esta conquista, sei que, pelo tempo em que me acompanhou e por sempre acreditar no meu potencial, consegui enchê-la de orgulho.

Aos meus amigos João, Caio e Dario, que estiveram comigo desde o início da faculdade, sempre tornando as noites de estudo mais leves e ajudando a enfrentar os desafios acadêmicos com bom humor e parceria. Mais do que colegas de curso, vocês se tornaram irmãos que levarei para a vida.

Por fim, sou imensamente grato aos professores que me auxiliaram nessa trajetória, não apenas no âmbito acadêmico, mas também no meu crescimento profissional. Meu sincero agradecimento ao Professor Alyson, à Professora Micila, à Professora Ana Lúcia e ao Professor Anderson, por toda dedicação e ensinamentos que foram essenciais para minha formação.

A todos vocês, meu mais profundo agradecimento!

RESUMO

Este trabalho propõe o estudo de métodos que busca unir conceitos da sociedade 5.0 e indústria 4.0 para serem utilizados a favor da sustentabilidade. Com o aumento da produção de resíduos plásticos ao redor do mundo e baixo índice de reciclagem do mesmo, foi observada a necessidade de amenizar o impacto no ecossistema. A metodologia usada quanto a natureza, foi uma pesquisa aplicada, explicativa, com delineado tecnológico. Como resultado, três métodos foram desenvolvidos para o reconhecimento de resíduos recicláveis. Por fim, o método selecionado utilizou uma rede neural convolucional para a criação e treinamento de um classificador destinado a uma máquina, acionada por tampas plásticas e metálicas. O objetivo desse sistema é promover a sustentabilidade e reduzir a quantidade de materiais recicláveis descartados de forma inadequada no meio ambiente.

Sendo assim, este trabalho busca realizar o desenvolvimento de classificadores de baixo custo, e, dentre eles, selecionar o melhor para ser utilizado em uma máquina movida a resíduos recicláveis.

Palavras-chave: visão computacional, resíduos recicláveis, classificador.

ABSTRACT

This work proposes the study of methods that seek to unite concepts from society 5.0 and industry 4.0 to be used in favor of sustainability. With the increase in the production of plastic waste around the world and its low recycling rate, there was a need to reduce the impact on the ecosystem. The methodology used was applied, explanatory research with a technological design. As a result, three methods were developed for recognizing recyclable waste. Finally, the method selected used a convolutional neural network to create and train a classifier for a machine, powered by plastic and metal lids. The aim of this system is to promote sustainability and reduce the amount of recyclable materials disposed of inappropriately in the environment.

Therefore, this work aims to develop low-cost sorters and select the best one to be used in an machine powered by recyclable waste.

Keywords: computer vision, recyclable waste, classifier.

Sumário

1	Introdução	9
1.1	Justificativa	10
1.2	Objetivo	11
1.2.1	Objetivo Geral	11
1.2.1	Objetivo Específico	12
1.3	Metodologia	12
1.4	Estrutura do TCC	13
2	Referencial Teórico	14
2.1	Imagem Digital	14
2.2	Visão Computacional	16
2.3	Inteligência Artificial	18
2.4	Machine Learning	20
2.5	Redes Neurais Artificiais	22
2.5.1	Aprendizado Profundo	24
2.5.2	Redes Neurais Convolucionais	25
3	Trabalhos Anteriores Relacionados	27
4	Materiais e Método de Desenvolvimento	28
4.1	Materiais	28
4.1.1	Visual Studio Code	28
4.1.2	Python	29
4.1.3	LabelIMG	30
4.1.4	Darknet	31
4.2	Método	32
4.2.1	Prova de Conceito Captura de Imagem	32
4.2.2	Treinamento HaarCascade	36
4.2.3	Treinamento YOLO	40
4.2.4	Resultados e Discussão	44
5	Conclusão	49
6	Referências	50

1. INTRODUÇÃO

A busca por meios de produção mais eficientes e a introdução de tecnologias emergentes deu início à era da indústria 4.0. Também chamada de a 4ª Revolução Industrial, com a finalidade de fazer com que as indústrias se tornem inteligentes, conectadas e digitalizadas. De acordo com Lima Pinto (2019) há termos usados de maneira globalizada na indústria 4.0 que expressam algumas tecnologias para automação e troca de dados como: Inteligência Artificial, Internet das Coisas (IoT), Computação em nuvem, Big Data e Sistemas Cyber-físicos.

A partir da evolução das tecnologias digitais e inovações propostas pela Indústria 4.0, o Japão, conforme Merluzzi (2018) cria um novo conceito para a sociedade que se beneficia dessas tecnologias, assim surge a “Super Smart Society 5.0”, conhecida como sociedade 5.0. Esse novo modelo de sociedade tem sido abordado desde 2016 pelos japoneses, e tem como propósito melhorar a qualidade de vida e atenuar problemas sociais, por exemplo: soluções para longevidade, cura de doenças extremas, mobilidade personalizada, previsões de catástrofes ambientais e cuidados com o meio ambiente.

Dessa forma a Indústria 4.0 e a nova sociedade “Smart” (Sociedade 5.0) propõe o uso de tecnologias a serviço da humanidade, para tornar a vida das pessoas mais adaptável, resiliente e sustentável. Este conceito está alinhado à sustentabilidade, uma vez que esta, conforme Sachs (2008) possui 08 dimensões (ecológica, econômica, social, cultura, psicológica, territorial, política nacional e internacional). Essas dimensões se entrelaçam a serviço da humanidade e não o contrário.

Neste sentido, este projeto de conclusão de curso busca integrar as metodologias tradicionais de automação às estratégias de tecnologias convergentes, onde a máquina é capaz de aprender com suas experiências e assim realizar tomadas de decisões. Assim, este trabalho busca a aplicação de modelos inteligentes integrados a um sistema de automação para classificação inteligente de resíduos sólidos recicláveis.

Dentro desta perspectiva, esta proposta busca o refinamento da codificação de algoritmos complexos e banco de dados para a mineração de informações, a fim de possibilitar uma resposta ao sistema causando um efeito na interação com o usuário e uma melhor usabilidade do dispositivo de automação.

Este trabalho foca no desenvolvimento de um modelo de inteligência artificial para o reconhecimento de resíduos recicláveis, integrando-se a um projeto mais amplo que também abrange software e mecânica elétrica. O objetivo final desse conjunto de esforços é projetar e construir uma máquina gamificada acionada por resíduos recicláveis. Essa iniciativa visa incentivar a reciclagem, reduzir a taxa de desperdício e aumentar os índices de reaproveitamento de materiais. O modelo desenvolvido neste trabalho desempenha um papel fundamental na identificação e classificação dos resíduos, contribuindo diretamente para a funcionalidade da máquina e para a promoção da sustentabilidade, alinhada à construção de um mundo mais sustentável.

1.1 JUSTIFICATIVA

Segundo um levantamento realizado pela WWF (World Wide Fund for Nature) no ano de 2018, o Brasil ocupou a quarta posição no ranking dos países que mais produzem resíduos plásticos no mundo, com cerca de 11 milhões de toneladas por ano. Não bastando estar em quarto lugar, o país ocupa uma posição no ranking, entre os países que menos reciclam seus resíduos plásticos. O Brasil recicla cerca de 1,28% de todo resíduo plástico gerado, e isso refere-se a 145 mil toneladas, índice bem abaixo da média quando comparado com outros países. Com o baixo índice de reciclagem, os resíduos que não passam por esse tratamento acabam trazendo consequências graves para nosso ecossistema, como: poluição do ar, poluição dos lençóis freáticos, intoxicação de animais, dentre outros.

De acordo com uma pesquisa realizada pela Associação Brasileira de Empresas de Limpeza Pública e Resíduos Especiais (Abrelpe, 2022) diz que aproximadamente um quarto das cidades brasileiras não conta com coleta seletiva, ou seja, isso representa cerca de 1.400 municípios sem políticas

públicas que promovam ou incentivem a separação do lixo reciclável pela população. Outra pesquisa conduzida pela Abrelpe em parceria com o Ibope revelou que mais de 70% dos brasileiros não separam o lixo comum do reciclável. Essa prática inadequada resulta em graves impactos ambientais, como a contaminação do solo e da água, a emissão de gases de efeito estufa, a redução da vida útil dos aterros sanitários e outros efeitos prejudiciais que comprometem diretamente a qualidade de vida da população brasileira.

Neste sentido, o desenvolvimento de abordagens tecnológicas para o tratamento desse tipo de resíduo é justificável, pois possibilita que ferramentas e conceitos vitais da Indústria 4.0 sejam aplicados à sustentabilidade e meio ambiente, possibilitando que aplicações desse tipo causem impactos importantes na vida das pessoas, principalmente voltados a redução do impacto causado pela falta da reciclagem de resíduos plásticos.

Além disso, este projeto de conclusão de curso justifica-se na aplicação de conceitos da sociedade 5.0 e tecnologias emergentes da indústria 4.0 para o desenvolvimento de um modelo que seja capaz de realizar o reconhecimento de resíduos através de visão computacional contendo o baixo custo possibilitando a replicação exponencial do dispositivo.

Tendo em vista que os projetos desenvolvidos na área de automação, ainda ocorrem de forma tradicional, com aplicações de automação simplificadas, onde um hardware tem um firmware executa uma tarefa específica a qual foi planejada, este projeto busca aplicar componentes de inteligência artificial possibilitando contribuição relevante para alavancar o desenvolvimento de automações inteligentes.

1.2 OBJETIVO

1.2.1 OBJETIVO GERAL

O objetivo geral deste trabalho é desenvolver um modelo inteligente para a análise automatizada de dados de resíduos descartados, utilizando métodos de visão computacional. O sistema será capaz de processar imagens capturadas por uma câmera, realizando a identificação e classificação dos resíduos de forma eficiente e precisa. Este modelo visa otimizar o processo de análise, substituindo

métodos manuais por uma solução automatizada e de baixo custo, contribuindo para a sustentabilidade e o aprimoramento de sistemas de reciclagem.

1.2.2 OBJETIVOS ESPECÍFICOS

Com o intuito de alcançar os objetivos propostos neste trabalho, foram definidos os seguintes objetivos específicos:

1. Criar um dataset robusto com resíduos recicláveis que busca-se detectar.
2. Desenvolver algoritmos utilizando técnicas de mineração e tratamento de dados para estruturar o dataset de forma adequada.
3. Aplicar técnicas de visão computacional para o reconhecimento de resíduos recicláveis.
4. Treinar modelos de aprendizado de máquina empregando métodos de processamento de imagens, normalização de dados e redes neurais profundas.
5. Comparar, selecionar e validar um dos modelos desenvolvidos com taxa de acerto superior a 90% para realizar a classificação dos resíduos inseridos na máquina.

1.3 METODOLOGIA

A metodologia usada quanto a natureza deste trabalho, foi baseada em uma pesquisa geral aplicada, explicativa, com delineamento tecnológico combinando abordagens qualitativas e quantitativas. A pesquisa bibliográfica será realizada para realizar o embasamento do estudo, realizando o levantamento de artigos, teses, dissertações e trabalhos relacionados a reconhecimento de objetos que possam nortear um direcionamento para o embasamento teórico e a definição da implementação do projeto utilizando metodologias como redes neurais e aprendizado de máquina para o treinamento de um modelo que possa realizar o reconhecimento de resíduos recicláveis. A abordagem qualitativa tange a respeito dos resultados alcançados pelo trabalho, onde será realizado uma análise e discussão abordando as principais limitações e desafios

De acordo com Yin (2015), a combinação de abordagens qualitativas e quantitativas pode ser fundamental para a compreensão de um fenômeno complexo, permitindo uma análise mais completa e detalhada dos resultados obtidos.

1.4 ESTRUTURA DO TCC

A organização deste trabalho está estruturada da seguinte forma:

A organização deste trabalho está estruturada da seguinte forma:

No Capítulo 1, será apresentada a introdução do trabalho, juntamente com a justificativa que embasa seu desenvolvimento. Além disso, serão expostos os objetivos gerais e específicos, seguidos da metodologia utilizada e da estrutura do TCC.

No Capítulo 2 será apresentado a fundamentação teórica necessária para o desenvolvimento do estudo, abordando conceitos relacionados a imagens digitais e visão computacional, inteligência artificial e técnicas de classificação de imagens.

No Capítulo 3, serão detalhados os trabalhos relacionados que serviram como base de conhecimento para o desenvolvimento deste estudo.

No Capítulo 4, serão apresentados os materiais utilizados no desenvolvimento do trabalho, bem como os métodos adotados para cada tipo de classificador desenvolvido.

O Capítulo 5 apresenta as conclusões finais do estudo e os trabalhos futuros. E por fim, no capítulo 6 as referências utilizadas para o desenvolvimento deste trabalho.

2. REFERENCIAL TEÓRICO

O objetivo deste capítulo é apresentar a base conceitual, científica e técnica que fundamentará o trabalho, servindo como alicerce para as discussões práticas desenvolvidas ao longo do TCC.

2.1 IMAGEM DIGITAL

Uma imagem digital é uma representação bidimensional de informações visuais em formato numérico que pode ser processada e analisada por sistemas computacionais contendo duas informações principais a intensidade e sua cor. Antonio Scuri (1999), diz que, em processamento de imagens usa-se o modelo matricial e a computação gráfica geralmente se baseia no modelo de objetos vetoriais, ele conclui que, nesses modelos os objetos são armazenados apenas a partir da descrição das coordenadas de seus vértices, sejam elas espaciais ou planares. Em termos simples, é uma matriz de valores que representam os níveis de intensidade ou cores dos pixels que compõem a imagem.

De acordo com Ogê Filho e Hugo Neto (1999), uma imagem monocromática pode ser matematicamente descrita por uma função $f(x,y)$, que representa a intensidade luminosa. O valor dessa função em qualquer ponto com coordenadas espaciais (x,y) é proporcional ao brilho da imagem naquele ponto, refletindo as variações de luminosidade ao longo da imagem. O X representa as linhas enquanto o Y representa as colunas, através de uma função de $f(x,y)$ é possível extrair as informações do pixel selecionado, informações como cor e intensidade..

Ao explorar uma os endereços de uma imagem, é possível encontrar os pixels, um pixel é o nome denominado a menor unidade de uma imagem digital. Cada pixel contém informações que representam a intensidade de luz, no caso de imagens em escala de cinza, ou uma combinação de cores, no caso de imagens coloridas. Em imagens coloridas, essa combinação é geralmente representada no espaço de cores RGB, onde cada pixel é composto por valores correspondentes às intensidades de vermelho, verde e azul.

Os canais de uma imagem representam as diferentes dimensões ou componentes que compõem os dados de cor ou intensidade de cada pixel.

Segundo Weeks (1996), em casos simples, cada pixel pode conter um único valor numérico que representa o nível de sinal em um ponto da imagem. Esse é o caso de imagens em escala de cinza, onde a intensidade varia de 0 a 255, com 0 representando preto e 255 representando branco. Já em imagens coloridas, cada pixel é formado por múltiplos valores numéricos, como no modelo RGB (Red, Green, Blue), no qual cada canal varia de 0 a 255. Nesse caso, os três canais combinam-se para representar uma ampla gama de cores.

Figura 1 : Imagem Digital



Fonte – Autor, 2024

Acima, foi criada uma imagem para ilustrar o conceito explicado sobre imagens. É possível observar uma representação no formato RGB, em que cada pixel é definido como uma função das coordenadas X e Y . Em cada coordenada, os valores dos canais R (vermelho), G (verde) e B (azul) variam de acordo com a cor e a intensidade presentes na imagem.

2.2 VISÃO COMPUTACIONAL

A Visão Computacional, um novo ramo da Ciência da Computação, é a forma que o computador enxerga o mundo, ou seja, é a forma que uma máquina consegue extrair informações de uma imagem. De acordo com Ballard e Dana (1987), a visão computacional é a construção de descrições explícitas e significativas de objetos físicos a partir de imagens buscando automatizar e integrar uma ampla gama de processos e representações utilizados na percepção visual.

Atreladas a dois pilares fundamentais da Indústria 4.0 – robótica e inteligência artificial –, atividades como inspeções de qualidade, automação de processos, segurança operacional e manutenção preditiva têm sido cada vez mais automatizadas. Nesse contexto, robôs equipados com câmeras podem ser treinados para interpretar imagens e tomar decisões de forma autônoma. De acordo com dados da Mordor Intelligence (2024), o mercado de visão computacional movimentava atualmente USD 17,25 bilhões e está projetado para dobrar, atingindo USD 39 bilhões até 2029. Esse crescimento reflete o reconhecimento das empresas sobre o potencial desse segmento para gerar negócios, automatizar tarefas, aumentar lucros e reduzir custos.

Ao abordar a visão computacional, Weeks (1996) apresenta uma sequência de etapas fundamentais para o processamento e tratamento de imagens. Essas etapas são:

1. **Aquisição da Imagem:** Este procedimento descreve como será realizada a aquisição de uma imagem, destacando dois elementos fundamentais em um sistema de visão computacional: o hardware e o software. O hardware refere-se aos equipamentos que compõem o ambiente, como câmeras, sensores e scanners, responsáveis pela captura dos dados visuais. Já o software desempenha o papel de processar essas imagens capturadas, transformando os dados brutos em informações úteis para análise e interpretação;
2. **Pré-Processamento:** Técnicas de pré-processamento são aplicadas nesta etapa com o objetivo de minimizar falhas

decorrentes do processo de aquisição da imagem. Essas técnicas visam melhorar a qualidade da imagem ao corrigir problemas relacionados à iluminação, contraste, distorções e nitidez, garantindo uma base mais consistente e precisa para as etapas subsequentes de análise.

3. **Segmentação:** Nesta etapa, a imagem é dividida em várias partes ou regiões com o objetivo de identificar e isolar objetos ou áreas de interesse. O principal propósito é agrupar pixels com características semelhantes, como cor, intensidade, textura ou localização, diferenciando-os das demais regiões da imagem.
4. **Identificação do Objeto:** Com a imagem devidamente pré-processada e segmentada, a identificação da região de interesse torna-se significativamente mais precisa e eficiente. Nessa etapa, as informações e características relevantes da imagem são extraídas, permitindo a análise detalhada e facilitando a identificação do objeto de forma confiável. Esse processo é essencial para garantir que os dados extraídos sejam representativos e úteis para as etapas subsequentes, como reconhecimento de padrões ou classificação.
5. **Reconhecimento de Padrões:** O reconhecimento de padrões é a etapa final no processamento de imagens, onde as informações extraídas nas fases anteriores, como características e regiões de interesse, são analisadas para identificar objetos, formas ou padrões específicos. Com a imagem previamente pré-processada, segmentada e suas características isoladas, é possível aplicar técnicas que associam essas informações a classes ou categorias previamente definidas.

O desenvolvimento do projeto atual seguiu o passo a passo descrito por Weeks, garantindo a aplicação estruturada de cada fase.

2.3 INTELIGÊNCIA ARTIFICIAL

De acordo com Russel e Norvig (2016) a Inteligência Artificial (IA) pode ser definida como "o ramo da ciência da computação que busca desenvolver sistemas computacionais capazes de realizar tarefas que normalmente exigiriam inteligência humana, como aprendizado, raciocínio, percepção e interação com o ambiente". Já McCarthy (2007) define inteligência artificial como a ciência e engenharia de criar máquinas inteligentes, especialmente programas de computador, capazes de realizar tarefas que, quando realizadas por humanos requerem inteligência. Apesar de serem autores distintos, suas definições convergem ao destacar que, de forma geral, a inteligência artificial é uma área da ciência que busca simular o raciocínio humano, permitindo que máquinas executem tarefas complexas de maneira inteligente e autônoma

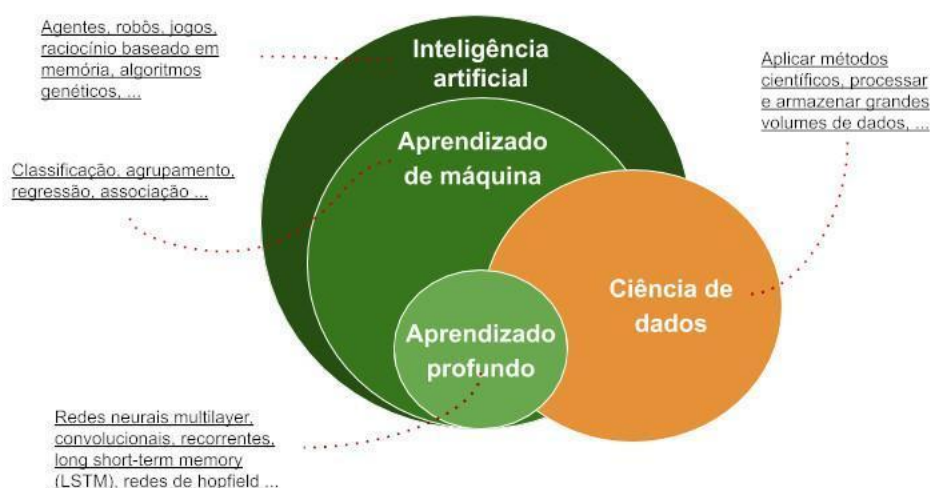
A inteligência artificial possui diferentes linhas de pensamento onde suas definições variam de autor para autor. Em um trabalho publicado por Dennis Gomes (2010) salienta que a Inteligência Artificial seguiu quatro principais linhas de pensamento, são elas:

1. Sistemas que pensam como seres humanos: "O novo e interessante esforço para fazer os computadores pensarem... máquinas com mentes, no sentido total e literal". (HAUGELAND, 1985).
2. Sistemas que atuam como seres humanos: "A arte de criar máquinas que executam funções que exigem inteligência quando executadas por pessoas". (KURZWEIL, 1990).
3. Sistemas que pensam racionalmente: "O estudo das faculdades mentais pelo seu uso de modelos computacionais". (CHARNIAK, MCDERMOTT, 1985).
4. Sistemas que atuam racionalmente: "A inteligência computacional é o estudo do projeto de agentes inteligentes". (POOLE et al., 1998).

Conforme descrito nas linhas de pensamento, pensar como um ser humano é diferente de atuar como um ser humano. Quando o contexto é pensar

como um ser humano refere-se a uma inteligência artificial para replicar ou simular os processos mentais humanos, como raciocínio, aprendizado, tomada de decisão e resolução de problemas. Enquanto atuar como seres humanos está relacionado a inteligência artificial tentar imitar a forma como os humanos reagem em situações específicas, sem necessariamente simular os processos cognitivos subjacentes. Newell e Simon (1976), “A cognição humana é essencialmente um processo computacional que pode ser representado por sistemas simbólicos”. Os autores fundamentam a ideia de que pensar como um humano está relacionado a representar processos cognitivos com sistemas computacionais.

Figura 2 - Áreas da Inteligência Artificial.



Fonte - SERPRO (<https://www.serpro.gov.br/menu/noticias/noticias-2019/democratizando-a-inteligencia-artificial>, acesso em julho de 2023).

A figura apresentada ilustra a relação entre as áreas da Inteligência Artificial (IA), Aprendizado de Máquina (Machine Learning), Aprendizado Profundo (Deep Learning) e Ciência de Dados, evidenciando suas interconexões e especificidades. No núcleo da representação, observa-se o Aprendizado Profundo, uma subárea especializada dentro do Aprendizado de Máquina, que utiliza técnicas avançadas como redes neurais convolucionais, redes recorrentes e LSTMs. Essa camada reflete a capacidade de modelos computacionais de lidar com problemas complexos em grandes volumes de dados.

2.4 MACHINE LEARNING

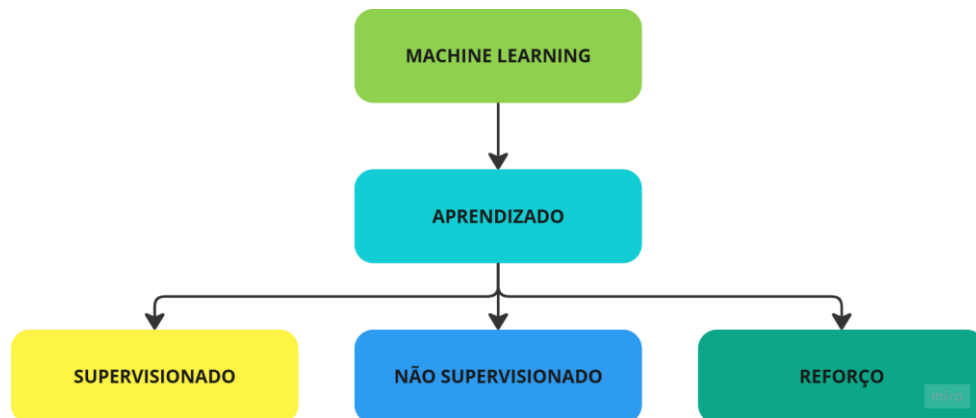
O subcampo de Machine Learning pode ser definido como um dos campos de estudo de maior importância para a ciência da computação, bem como para a computação aplicada em si. Este subcampo se relaciona com o campo de pesquisa da Inteligência Artificial, o qual possui suas raízes no desenvolvimento de modelos que possam simular a habilidade humana para solução de determinada tarefa, e até mesmo na relação intrínseca entre seres humanos e máquinas (TURING, 1950).

Mitchell (1997), define Machine Learning como um subcampo da inteligência artificial que se dedica ao desenvolvimento de métodos e algoritmos capazes de aprender padrões a partir de dados. Em vez de seguir apenas regras definidas explicitamente por programadores, os modelos de aprendizado de máquina identificam correlações, regularizadas e estruturas nos dados realizando o ajustes dos parâmetros de modo a generalizar o conhecimento adquirido para novos casos.

De acordo com Arantes Filho (2022), a questão do aprendizado de máquina pode ser levada ao entendimento do significado da inteligência de máquinas em si, através de modelos de aprendizado profundo que possam se aproximar da forma como o cérebro humano pode processar as informações e executar suas inferências. Ainda segundo o autor, o aprendizado de máquina pode ser definido como modelos e algoritmos que recebem como entrada um conjunto de dados e exemplos que estão relacionados a determinados fenômenos, no qual o objetivo é a solução de determinada tarefa ou problema, tais como a classificação de dados ou a regressão de dados (BISHOP, 2006; GERON, 2017; ARANTES FILHO, 2022).

Ao estudar técnicas de machine learning, identificam-se três tipos principais de métodos, cada um com suas especificidades e características próprias:

Figura 3 – Distribuição dos tipos de Machine Learning



Fonte – Autor, 2024

O aprendizado supervisionado é quando o algoritmo recebe um conjunto de exemplos de entrada acompanhados de suas respostas corretas. Isso permite que o modelo aprenda padrões para obter uma saída correta. De acordo com Mitchell (1997), essa abordagem é treinada a partir de dados rotulados, conjuntos de exemplos onde cada instância é acompanhada de um rótulo ou valor-alvo conhecido, seu objetivo é aprender uma função de mapeamento entre as variáveis de entrada e a saída desejada.

No aprendizado não supervisionado o algoritmo recebe apenas dados de entrada sem rótulos ou saídas previamente definidas. Diferente do aprendizado supervisionado, o aprendizado não supervisionado baseia-se exclusivamente na exploração de dados. O Resultado pode ser identificação de clusters, a descoberta de variáveis latentes, a redução de dimensionalidade ou a detecção de anomalias (Hastie, Tibshirani & Friedman, 2009).

Por último, o aprendizado por reforço. Um agente aprende a tomar decisões ao interagir com um ambiente, recebendo recompensas ou penalidades em resposta às suas ações. O Agente explora o ambiente, experimentando ações e observando as consequências de suas escolhas. Sutton & Barto (2018), definem como uma área inspirada na psicologia comportamental, na qual um agente aprende a agir em um ambiente de modo a maximizar uma recompensa cumulativa.

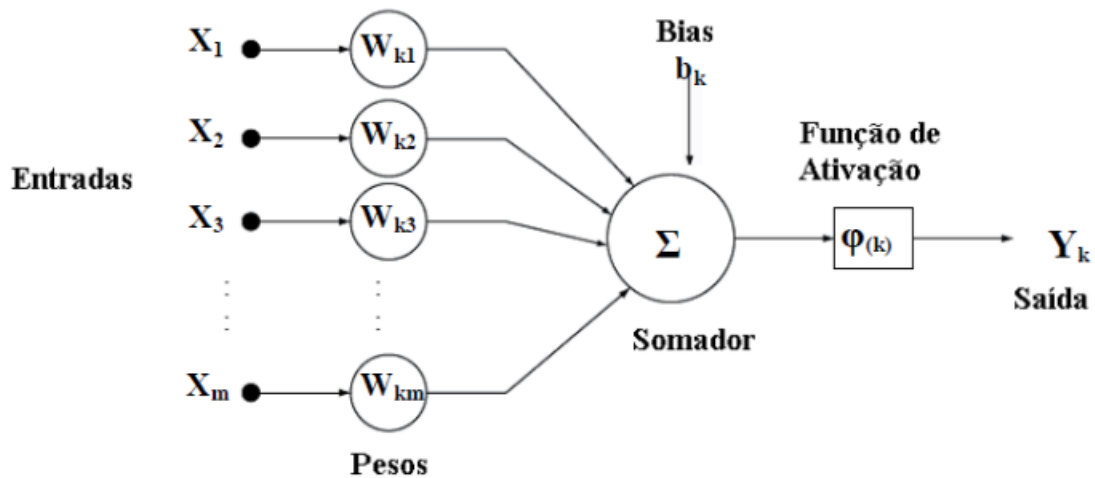
Diversas abordagens em machine learning têm sido desenvolvidas, tanto para classificação quanto para detecção de objetos. Os autores, De Espírito Santo e De Oliveira Filho (2020), por exemplo, desenvolveram um sistema para a detecção e contagem de bovinos em imagens aéreas utilizando visão computacional realizada por técnicas de processamento de imagens e visão computacional, a pesquisa refere-se à imagens tiradas de um drone, vista superior em relação ao gado, a uma altura de 30 a 50 metros e a partir daí são utilizados filtros para amenizar ruídos, conversão do sistema de cores, aplicação de processos de erosão e dilatação da imagem para ocorrer a detecção de bois.

2.5 REDES NEURAIS ARTIFICIAIS

A inteligência biológica surge a partir de entidades básicas, chamadas neurônios, interconectadas em redes permitindo a transmissão de informações. Daí surge a inteligência humana, onde, através de informações captadas por sensores, há o processamento dessa informação que gera uma ação. Tomas Walter, diz que, as redes neurais artificiais surgem como uma tentativa de copiar a estrutura e o funcionamento do cérebro em um ambiente técnico, realizando entendimento da inteligência residente nos neurônios e mapeá-la para uma estrutura artificial.

Em 1943, Warren McCulloch e Walter Pitts propuseram um modelo matemático do neurônio biológico, o que marcou o avanço no ramo da inteligência artificial. Esse modelo ficou conhecido como neurônio McCulloch-Pitts (M-P) sendo uma aproximação simplificada do neurônio real, o que possibilitou a construção de algoritmos de redes neurais. Esse modelo é baseado em uma combinação linear de entradas com seus respectivos pesos, seguido de uma função de ativação que determina a saída.

Figura 4 – Neurônio Artificial M-P

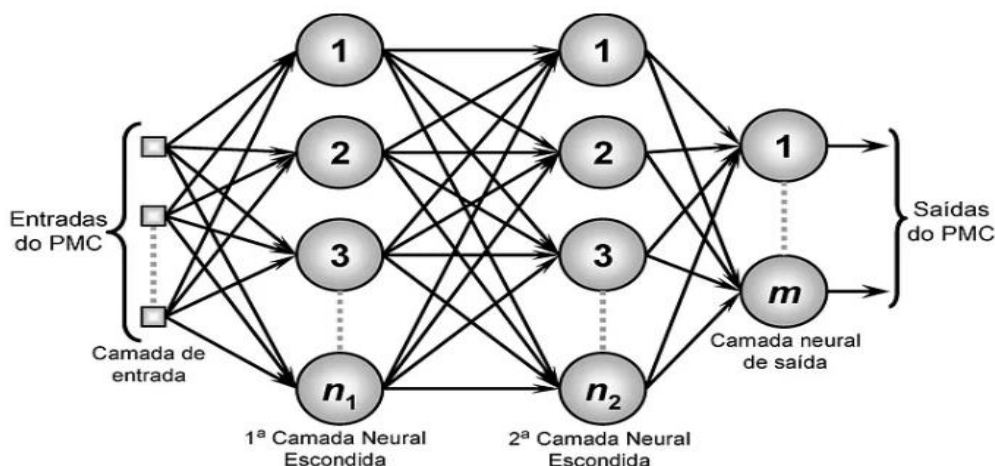


Fonte – Neurônio Artificial de McCulloch-Pitts (SOUZA, 2011)

Em 1958, Frank Rosenblatt apresentou uma das primeiras redes neurais artificiais, conhecida como Rede Perceptron. Baseada no modelo do neurônio de McCulloch-Pitts, com a inclusão de uma regra de aprendizagem, o que possibilita à rede ajustar os pesos das conexões com base nos erros cometidos durante o processo de treinamento. Através dessa inclusão ela pôde aprender a partir de exemplos e generalizar padrões em diferentes contextos. A estrutura desse modelo obedece os seguintes itens:

1. Entradas: Binários de entrada que podem assumir valores que variam de 0 ou 1;
2. Pesos: Cada entrada possui um peso associado, representando a importância relativa de cada entrada;
3. Somador: As entradas são multiplicadas por seus respectivos pesos e somadas;
4. Função de Ativação: O valor resultante é comparado a um limiar; Se o valor da soma for maior ou igual ao limiar, o neurônio dispara saída = 1, caso contrário, saída = 0;

Figura 5 – Rede Perceptron



Fonte – (<https://medium.com/ensina-ai/rede-neural-perceptron-multicamadas-f9de8471f1a9>, acesso em Dezembro de 2024)

2.5.1 APRENDIZADO PROFUNDO

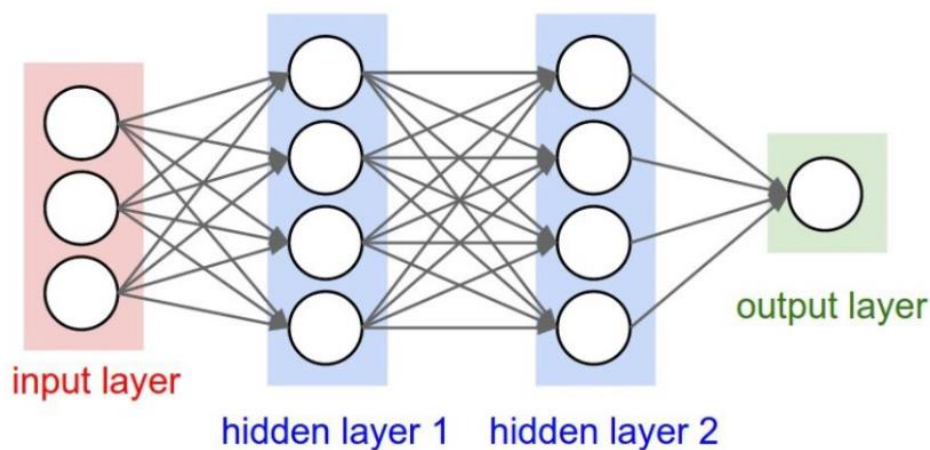
O Aprendizado profundo, também conhecido como Deep Learning, é um subcampo da inteligência artificial e do aprendizado de máquina que utiliza redes neurais profundas para aprender padrões complexos a partir de grandes volumes de dados. Ele é inspirado no funcionamento do cérebro humano e se baseia em redes neurais artificiais com múltiplas camadas. Segundo LeCun, Bengio e Hinton (2015), o aprendizado profundo abrange o desenvolvimento de modelos compostos por muitas camadas de neurônios artificiais, onde cada camada processa o output da camada anterior, permitindo a extração de características hierárquicas dos dados. Isso resulta em uma capacidade aprimorada de modelar padrões complexos e realizar inferências a partir de grandes volumes de dados.

O Aprendizado Profundo utiliza redes neurais profundas, isto é, com várias camadas ocultas que realizam transformações sucessivas nos dados de entrada até obter uma saída desejada. As principais camadas de uma rede neural incluem: 1. Camada de Entrada, também conhecida como Input Layer, nela os dados brutos são recebidos, como imagens, texto ou áudio; 2. Camadas Ocultas (Hidden Layers), essas camadas realizam os cálculos matemáticos para extrair características e aprender padrões; 3. Camada de Saída, Output Layer),

que são as camadas que geram o resultado final, podendo ser uma classificação, uma previsão ou a geração de dados.

Grande parte das redes são treinadas usando backpropagation e otimização por gradiente descendente, ajustando pesos dos neurônios para minimizar erros. Na imagem abaixo é possível identificar as camadas conforme apresentadas acima.

Figura 6 - Redes Neurais.



Fonte-(<https://getthematic.com/insights/what-is-deep-learning/>, acesso em Dezembro de 2024).

Algumas arquiteturas comuns utilizadas de Deep Learning são: Redes neurais recorrentes (RNNs), usadas para dados sequenciais, como séries temporais e linguagem natural; Redes generativas adversariais (GANs), essa rede é capaz de criar novos dados realistas como imagens e vídeos; Transformers, que são utilizadas em NLP, amplamente utilizada no GPT e DeepSeek e por último as Redes neurais convolucionais as CNNs, que são especializadas em imagens e visão computacional, sendo essa, a arquitetura abordada no trabalho.

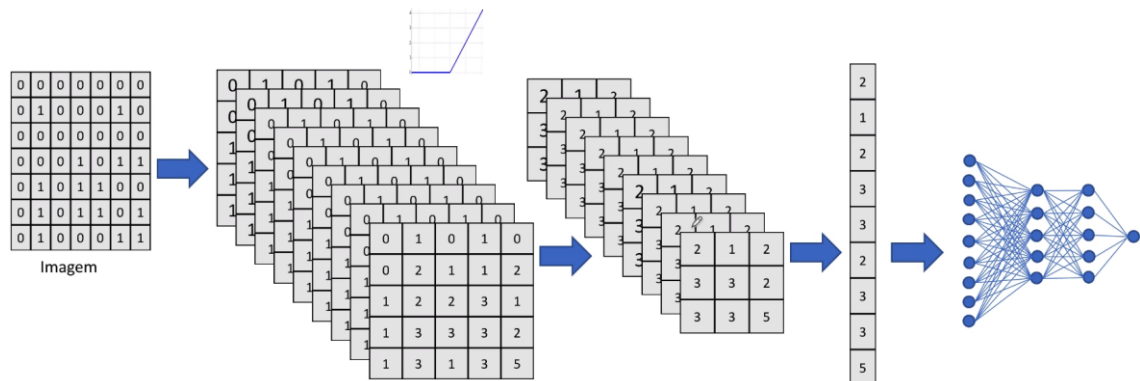
2.5.2 REDES NEURAS CONVOLUCIONAIS

As redes neurais convolucionais, CNNs, são um tipo de arquitetura de aprendizado profundo projetada para processar dados estruturados em forma de grade, como presente em imagens. Através de operações de convolução conseguem extrair automaticamente características dos dados. As redes neurais

convolucionais são uma arquitetura poderosa para análise de imagens, permitindo a extração automática de características hierárquicas (LECUN et al., 1998, p. 2559).

Figura 7 - Ilustração de uma Rede Neural Convolucional.

Rede neural convolucional



Fonte - IA Expert Academy (<https://iaexpert.academy/>, acessado em dezembro de 2024).

Uma rede neural convolucional é formada por várias camadas, sendo elas:

- Camada de Convolução (Convolutional Layer):** Aplica filtros sobre a imagem de entrada para extrair características, como bordas, texturas e formas;
- Camada de Pooling (Pooling Layer):** Reduz a dimensionalidade da saída da convolução, mantendo as informações mais relevantes.
- Camada de Normalização e Ativação:** Utiliza funções como ReLU (Rectified Linear Unit) para introduzir não-linearidade e normalizar os dados.
- Camadas totalmente conectadas (Fully Connected Layers):** Após a extração de características, os dados são achatados e passados para camadas densas que realizam a classificação final.

3. TRABALHOS ANTERIORES RELACIONADOS À DESENVOLVIMENTO DE METODOLOGIAS PARA RECONHECIMENTO DE RESÍDUOS RECICLÁVEIS

O objetivo desta seção é apresentar trabalhos e/ou publicações científicas similares ao presente trabalho, com intuito de sustentar a importância do tópico selecionado e bem como discutir os métodos utilizados por outros autores a fim de comparação com o método aqui elaborado e resultados obtidos.

3.1 - Lima et al. (2020), em seu trabalho intitulado como: Utilização de redes neurais convolucionais para a detecção de objetos em imagens aéreas adquiridas por drones; faz o uso de Redes Neurais Convolucionais (RNCs) para realizar a detecção de objetos em imagens aéreas capturadas por drones, com o objetivo de identificar tanques de água e outros reservatórios que podem servir como criadouros do mosquito *Aedes aegypti*. Átraves da dificuldade de realizar essa verificação de forma manual, esse estudo surge como uma resposta, verificando de forma automatizada através das imagens adquiridas por Veículos Aéreos Não Tripulados (VANTs), especialmente em regiões urbanas com maior incidência de doenças transmitidas por este mosquito.

A metodologia empregou um conjunto de 142 imagens capturadas por um drone DJI Phantom 3 Professional, com resolução de 3000x4000 pixels. Para o treinamento da RNC, foram utilizadas 2.498 subimagens de 200x200 pixels, nas quais técnicas de data augmentation foram aplicadas para expandir o conjunto de dados. A arquitetura desenvolvida contou com duas camadas convolucionais, duas camadas de pooling e duas camadas totalmente conectadas, utilizando a API Keras sobre a plataforma TensorFlow.

Os resultados obtidos mostraram uma acuracidade de 93,37% para a validação do treinamento e uma acuracidade de 79,74% na etapa de testes, mostrando que a abordagem proposta possui grande potencial para aplicações reais. Apesar dos bons resultados, os autores notaram a presença de falsos positivos e negativos, atribuídos às limitações do conjunto de dados e à necessidade de ajustes adicionais no treinamento. O estudo também sugere futuras extensões para a detecção de cenários compostos por pequenos objetos acumuladores de água, como lixo a céu aberto e calhas obstruídas (LIMA et al., 2020).

4. MATERIAIS E MÉTODO DE DESENVOLVIMENTO

O presente capítulo relata os materiais, softwares, sites e ferramentas utilizadas para o desenvolvimento do trabalho. Posteriormente, o método abordado é discutido amplamente, bem como é feito um resumo e apresentação das arquiteturas utilizadas para a construção de cada modelo individualmente.

4.1 MATERIAIS

4.1.1 VISUAL STUDIO CODE

O Visual Studio Code é um editor de código-fonte desenvolvido pela Microsoft e lançado em abril de 2015. Gratuito e multiplataforma, ele está disponível para Windows, macOS e Linux. Este software permite o desenvolvimento em diversas linguagens de programação, como C, C++, Python, HTML/CSS, entre outras. Sua gratuidade e a ampla comunidade de usuários garantem uma vasta biblioteca de extensões e plugins, que aprimoram a experiência de programação, tornando-a mais eficiente e personalizada.

A utilização dessa ferramenta foi fundamental para o desenvolvimento deste trabalho. Por meio dela, foi possível criar scripts, realizar o treinamento de modelos e conduzir testes nas aplicações desenvolvidas.

Figura 8 - Visual Studio Code.



Fonte-(<https://www.stickpng.com/pt-br/img/icones-logos-emojis/empresas-tecnicas/logotipo-completo-do-visual-studio-code>, acesso em Dezembro de 2024).

4.1.2 PYTHON

Python é uma linguagem de programação criada por Guido Van Rossum em 1990, no Instituto Nacional de Pesquisa para Matemática e Ciência da Computação da Holanda. Considerada uma linguagem de alto nível, ela se destaca pela sua sintaxe simples se assemelhando à língua inglesa, o que facilita a leitura e compreensão de programas. Ao comparar programas escritos em outras linguagens de programação, a escrita em python possui menos linhas de código. Luiz Borges (2014), completa dizendo que, a Linguagem de Programação Python inclui diversas estruturas de alto nível (listas ,dicionários, data/hora, complexos e outras) e uma vasta coleção de módulos prontos para uso, além de frameworks de terceiros que podem ser adicionados.

Além de sua facilidade de aprendizado e recursos nativos, a linguagem Python conta com um vasto conjunto de ferramentas voltadas para o campo da Inteligência Artificial e Machine Learning. Exemplos notáveis incluem as bibliotecas PyTorch e TensorFlow, amplamente utilizadas para o desenvolvimento e treinamento de modelos complexos.

No contexto da Inteligência Artificial, onde o treinamento de modelos exige o processamento de grandes volumes de dados, o Python se destaca por oferecer bibliotecas robustas que abrangem todo o fluxo de trabalho, desde o tratamento e preparação dos dados até a implementação dos algoritmos de treinamento.

De acordo com um estudo realizado em 2019 pela KDnuggets, 1800 Engenheiros de softwares participaram do estudo, a linguagem de programação Python obteve a primeira colocação como a linguagem mais amplamente utilizada em ciência de dados, aprendizado de máquina e plataformas de análise, ficando com, aproximadamente, 65,8% dos participantes.

Figura 9 - Logotipo da linguagem de programação Python.

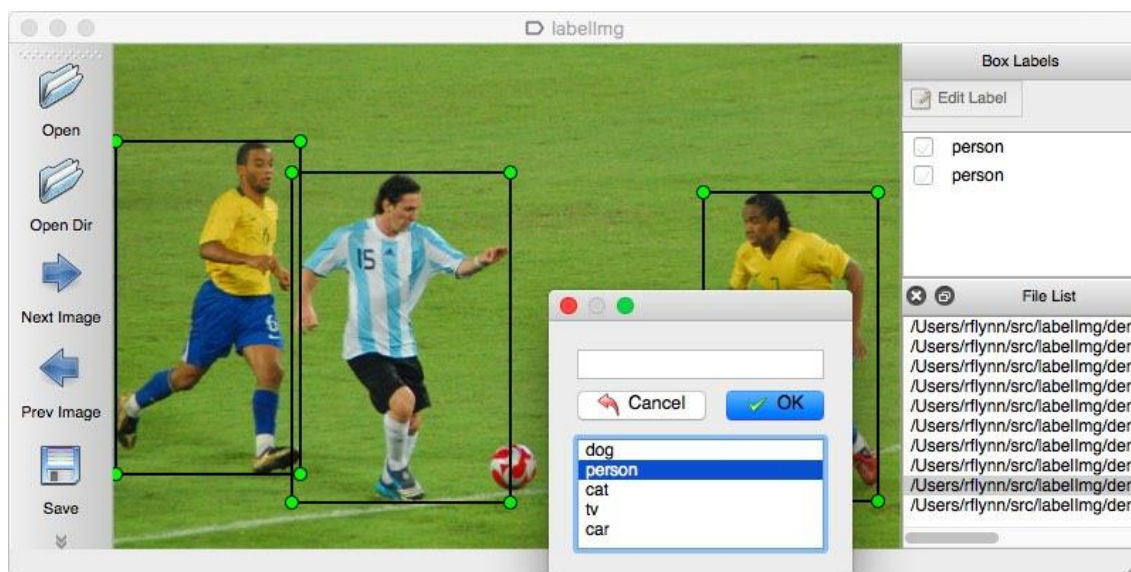


Fonte - Wikipedia (https://pt.m.wikipedia.org/wiki/Ficheiro:Python_logo_and_wordmark.svg, acesso em setembro de 2024).

4.1.3 LABELIMG

O Labelimg é uma ferramenta de código aberto, escrita em Python e baseada no framework Qt. Ela é utilizada para realizar a anotação manual de imagens em projetos que utilizam visão computacional. Sua principal função é permitir que os usuários desenhem caixas delimitadoras ao redor de objetos de interesse, gerando arquivos de anotação nos formatos PASCAL VOC (.xml) e YOLO (.txt), que são amplamente utilizados para treinamento de modelos de detecção de objetos.

Figura 10 - Utilização LabelIMG.



Fonte - Github (<https://github.com/HumanSignal/labelimg>, Acesso em Agosto de 2024)

4.1.4 DARKNET

A Darknet é um framework de código aberto desenvolvido por Joseph Redmon, lançado em 2016, possui o objetivo de fornecer uma estrutura leve e eficiente para redes neurais convolucionais (CNNs), otimizadas para processamento em tempo real. Este framework foi desenvolvido em C e em CUDA, a Darknet é amplamente utilizada no treinamento e implementação da família de modelos YOLO (You Only Look Once), algoritmo bastante eficiente para detecção de objetos em imagens e vídeos, sendo utilizado no presente trabalho. Sua capacidade de operar em GPUs, aliada à sua estrutura modular e de fácil customização, torna a Darknet uma ferramenta essencial para realizar o desenvolvimento de modelos que utilizam visão computacional.

Por se tratar de um framework de código aberto, a Darknet permite que os usuários modifiquem e personalizem seu código-fonte, sendo altamente flexível e adaptável para diferentes tipos de aplicações.

Figura 11 - Logotipo da Darknet.



Fonte - Github (<https://github.com/frankzhangrui/Darknet-Yolo/blob/master/README.md>, acesso em agosto de 2024)

4.2 MÉTODO

O presente trabalho propõe a aplicação de visão computacional para a identificação de resíduos recicláveis em imagens capturadas por câmeras. Para o desenvolvimento, foram exploradas três abordagens principais, com o objetivo de selecionar aquela que apresentou os melhores resultados em termos de tempo de processamento e precisão na detecção. Nos parágrafos seguintes, cada metodologia será detalhada, incluindo a descrição de cada etapa do processo.

4.2.1 PROVA DE CONCEITO CAPTURA DE IMAGEM

Inicialmente, foi realizado um estudo para determinar a posição ideal da câmera, visando maximizar a precisão na detecção dos resíduos. Três posições diferentes foram testadas. Na primeira, a câmera foi posicionada paralelamente a um duto por onde o resíduo passa, permitindo a detecção no momento em

que a tampa atravessa o duto, conforme ilustrado na imagem pela prova de conceito (PoC) desenvolvida.

Figura 12 - POC 1 de Captura de Imagem.



Fonte – Autor, 2021

A partir da primeira posição de captura, foi possível observar que a câmera não atendia aos requisitos necessários para registrar imagens com o objeto em movimento de forma nítida. Como consequência, muitas imagens foram capturadas sem o objeto de interesse ou com ele desfocado, o que comprometeria significativamente o reconhecimento, mesmo com a aplicação de técnicas avançadas de processamento de imagens. Esse problema resultaria na perda de informações essenciais, dificultando a extração de características relevantes para a detecção precisa do resíduo.

Abaixo, são apresentadas imagens ilustrativas que evidenciam as limitações dessa abordagem, destacando os desafios encontrados, como o desfoque e a inconsistência na captura do objeto no momento exato de sua passagem pelo duto.

Figura 13 - POC 1 de Imagem Capturada.



Fonte – Autor, 2021

Na segunda posição testada, a câmera foi posicionada acima do recipiente de coleta dos resíduos, capturando as imagens a partir de uma visão superior. Essa configuração permitiu a análise dos resíduos vistos de cima, oferecendo uma perspectiva diferente para a detecção.

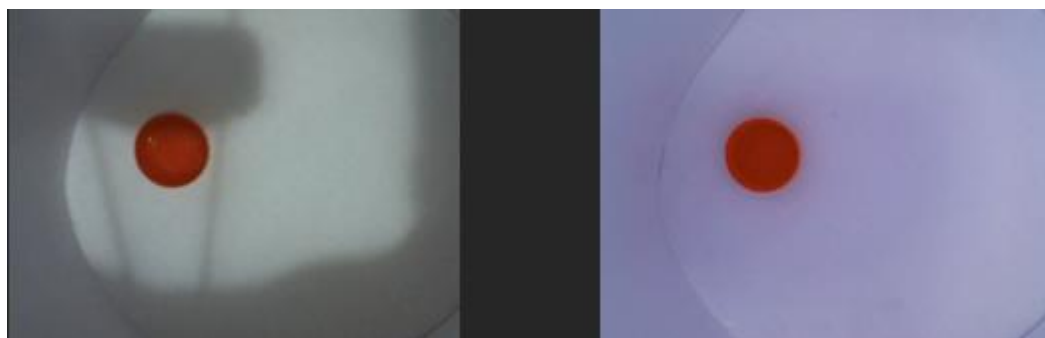
Figura 14 - POC 2 de Captura de Imagem.



Fonte – Autor, 2021

Como resultado dessa segunda prova de conceito (PoC), foi possível obter as seguintes capturas dos resíduos:

Figura 15 - POC 2 de Imagem Capturada.



Fonte – Autor, 2021

Por fim, na terceira posição testada, a câmera foi posicionada em um ângulo de 45° em relação ao recipiente onde os resíduos são depositados. Nesse cenário, o resíduo é inserido, passa pelo duto e, ao chegar ao recipiente, permanece alinhado com a câmera, que também está inclinada a 45°. Dessa forma, tanto a câmera quanto o resíduo mantêm uma orientação paralela entre si, formando um ângulo de 45° em relação ao solo.

Figura 16 - POC 3 de Captura de Imagem.



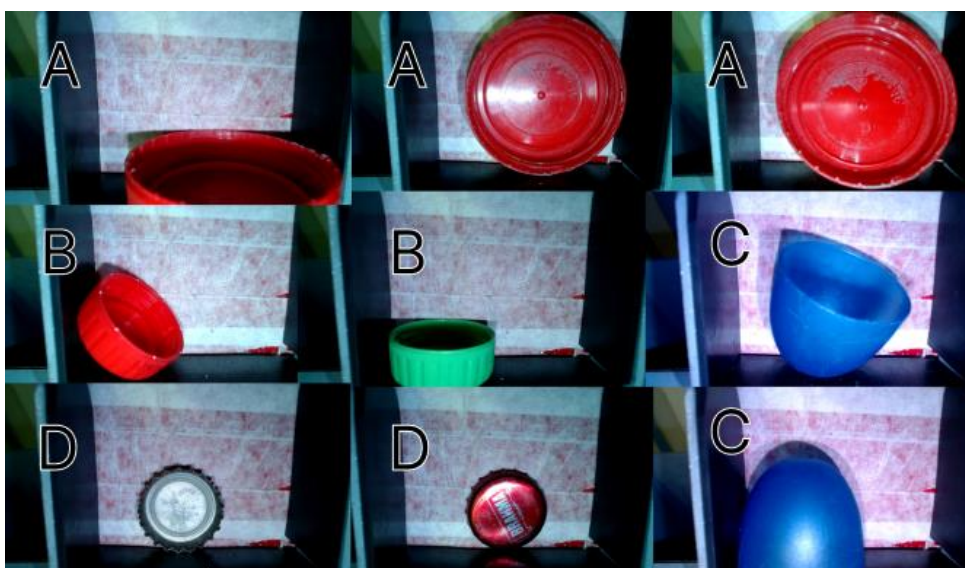
Fonte – Autor, 2021.

Como proposta inicial para a detecção de resíduos recicláveis, foram definidos os itens que seriam detectáveis e, conseqüentemente, utilizados no treinamento dos modelos. A seleção foi baseada na recorrência desses objetos

no descarte e na viabilidade de detecção por meio de visão computacional. Dessa forma, os seguintes itens foram escolhidos para compor o conjunto de dados e treinar os modelos:

- Tampas de garrafas PET;
- Tampas de garrafas de alumínio;
- Tampas de desodorantes Rexona;
- Tampas de fermento Oetker;

Figura 17 - Resíduos Recicláveis Selecionados.



A - Tampa de Ferment Oetker; B - Tampa de Garrafa Pet; C - Tampa de Desodorante Dove; D - Tampa de Cerveja.

Fonte – Autor, 2021

A escolha desses itens levou em consideração suas características visuais distintas, como formato, cor e textura, permitindo que o modelo aprendesse padrões específicos para diferenciá-los de outros objetos.

4.2.2 TREINAMENTO HAARCASCADE

O primeiro método adotado baseia-se no desenvolvimento de um modelo utilizando a abordagem Haar Cascade, uma técnica amplamente utilizada para detecção de padrões em imagens. O processo inicia-se com a coleta de imagens contendo o resíduo a ser identificado. Para essa coleta, foram utilizadas uma

câmera de celular e uma webcam, permitindo a captura de fotos dos objetos de interesse em diferentes ângulos, iluminações e distâncias.

Para iniciar o treinamento do modelo, foi necessário selecionar e classificar as imagens em dois grupos: imagens positivas e imagens negativas.

As imagens positivas são aquelas que contêm os resíduos que desejamos ensinar o modelo a reconhecer. Inicialmente, no primeiro treinamento, foram coletadas 10 fotos de diferentes resíduos. Para ampliar a amostragem e aumentar a diversidade dos exemplos, utilizou-se a ferramenta *createsamples* da biblioteca OpenCV, permitindo a geração de um conjunto expandido de imagens sintéticas. Essa abordagem foi fundamental para melhorar a capacidade do modelo de generalizar a detecção em diferentes condições.

Em seguida, foram analisadas as imagens negativas, que correspondem a imagens que não contêm os resíduos de interesse. Esse conjunto é essencial para o treinamento, pois ajuda o modelo a distinguir corretamente os objetos relevantes de outros elementos presentes no ambiente, reduzindo falsos positivos na detecção.

Figura 18 - Utilizadas para Treinamento.



Fonte – Autor, 2021.

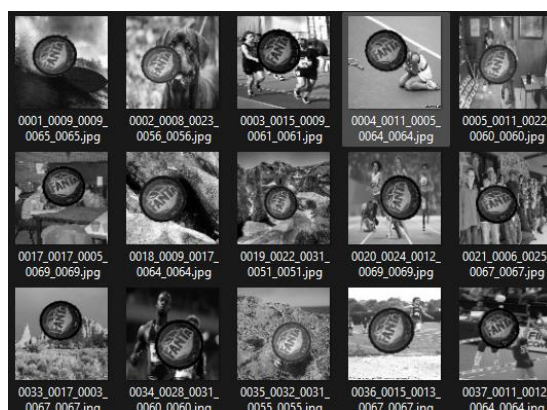
Após a coleta das imagens, é necessário realizar um pré-processamento antes de iniciar o treinamento do modelo. Esse processo é essencial para padronizar os dados e otimizar o desempenho do modelo.

O primeiro passo do pré-processamento é a conversão das imagens para escalas de cinza. Essa etapa é crucial, pois reduz a complexidade dos dados ao

eliminar informações de cor, permitindo que o modelo trabalhe com apenas um canal (kernel) em vez de três (RGB). Com isso, o processamento se torna mais eficiente, reduzindo o tempo de treinamento e o consumo de memória.

Em seguida, as imagens são redimensionadas para um tamanho padrão. Esse redimensionamento é fundamental para garantir que todas as imagens tenham dimensões uniformes, evitando inconsistências durante o treinamento. Além disso, reduzir o tamanho das imagens diminui a quantidade de pixels a serem processados, tornando o treinamento mais rápido sem comprometer a capacidade do modelo de aprender as características essenciais dos resíduos recicláveis.

Figura 19 - Resultado do CreateSamples



Fonte – Autor, 2021.

Em seguida, as imagens são redimensionadas para um tamanho padrão. Esse redimensionamento é fundamental para garantir que todas as imagens tenham dimensões uniformes, evitando inconsistências durante o treinamento. Além disso, reduzir o tamanho das imagens diminui a quantidade de pixels a serem processados, tornando o treinamento mais rápido sem comprometer a capacidade do modelo de aprender as características essenciais dos resíduos recicláveis.

Após a aquisição das imagens de interesse e a realização do pré-processamento, o treinamento do modelo pode ser iniciado. O processo de treinamento foi feito utilizando o script abaixo. Esse script inclui a definição dos parâmetros de treinamento.

Segue o código utilizado para o treinamento do modelo:

`opencv_traincascade -data classifier -vec samples.vec -bg negatives.txt -numStages 10 -minHitRate 0.995 -maxFalseAlarmRate 0.5 -numPos 1000 -numNeg 500 -w 24 -h 24 -mode ALL`

Tabela 1 - Lista de Comandos, Descrições e valores utilizados para treinamento do modelo.

Comando	Descrição	Valor
data classifier	Define o diretório onde o classificador treinado será armazenado	C:/
vec	Especifica o arquivo contendo as amostras positivas geradas pelo createsamples	samples.vec
-bg	Indica o arquivo com a lista de imagens negativas	negatives.txt
numStages	Define o número de estágios do classificador	10
minHitRate	Estabelece a taxa mínima de detecção desejada para cada estágio	0.995
maxFalseAlarmRate	Define a taxa máxima de falsos positivos permitida por estágio	0.5
numPos	Especifica o número de imagens positivas utilizadas no treinamento	1000
numNeg	Indica a quantidade de imagens negativas	500
-w -h	Define a dimensão das janelas de detecção	24 x 24
mode	Configura o treinamento para utilizar todas as técnicas disponíveis para otimização	ALL

Fonte – Autor, 2021.

Com base no primeiro modelo treinado, foram desenvolvidos mais três modelos adicionais com o objetivo de avaliar e comparar seu desempenho. A intenção foi testar a eficácia de cada versão, analisando a precisão na detecção

dos resíduos, bem como as taxas de falsos positivos e falsos negativos. Essa abordagem permitiu identificar qual modelo apresentava a melhor relação entre acurácia e confiabilidade, garantindo uma detecção mais eficiente e robusta.

Tabela 2 - Modelos Haar Cascade.

Nome	Imagens Positivas	Imagens Negativas	Estágios de Treinamento
Modelo 1	4000	3200	5
Modelo 2	4000	3200	10
Modelo 3	8000	4000	15

Fonte – Autor, 2021

Por meio do treinamento dos modelos utilizando a abordagem Haar Cascade, foi possível realizar a detecção de alguns resíduos plásticos com certo nível de precisão.

A imagem abaixo ilustra um exemplo da detecção realizada, evidenciando a eficácia do modelo treinado na identificação dos resíduos plásticos dentro do ambiente testado.

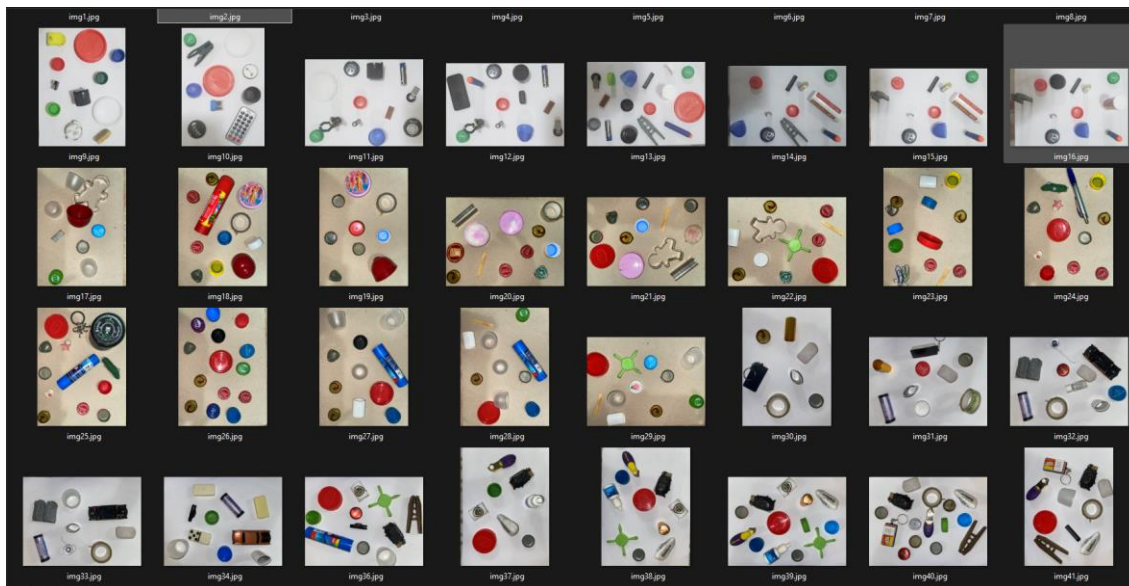
4.2.3 TREINAMENTO YOLO

Para o segundo treinamento do modelo, foi necessário aprimorar a qualidade do conjunto de dados, ampliando o número de imagens dos resíduos recicláveis que se busca-se detectar. Dessa forma, um novo dataset foi criado, contendo uma quantidade maior de amostras, incluindo variações de iluminação, ângulos e fundos distintos. Além disso, foram adicionadas imagens de objetos semelhantes aos resíduos de interesse, a fim de melhorar a capacidade do modelo de diferenciar corretamente os materiais recicláveis de outros elementos presentes no ambiente. Essa abordagem visou reduzir a taxa de falsos positivos e falsos negativos, tornando a detecção mais precisa e robusta.

Com esse refinamento do conjunto de dados, o modelo pôde aprender padrões mais complexos e generalizar melhor a identificação dos resíduos

recicláveis, garantindo um desempenho superior em relação ao treinamento inicial.

Figura 20- Novo DataSet



Fonte – Autor, 2021.

Com um novo dataset montado, o segundo método para realizar a detecção de resíduos utilizado foi o YOLO. O YOLO é um modelo baseado em redes neurais convolucionais (CNN) que permite a identificação de objetos em tempo real, sendo amplamente utilizado devido à sua alta precisão e eficiência computacional. O treinamento do modelo seguiu as etapas descritas a seguir.

Após a coleta do dataset, foi necessário organizar os arquivos em pastas separadas para armazenar as imagens e seus respectivos rótulos. Os rótulos são arquivos de texto associados a cada imagem, contendo as informações sobre as coordenadas do objeto a ser identificado. No formato utilizado pelo YOLO, cada imagem deve ter um rótulo correspondente, seguindo a estrutura:

<ID da Classe> <Coordenada X> <Coordenada Y> <Largura> <Altura>

Onde:

- ID da Classe: identifica a categoria do objeto na imagem.
- Coordenada X e Coordenada Y: representam o centro da bounding box, normalizados entre 0 e 1.

- Largura e Altura: definem o tamanho da bounding box, também normalizados.

A numeração do ID da Classe varia conforme a quantidade de objetos que se deseja detectar. No presente projeto, foram utilizadas duas classes:

- 1 - Cap (Tampas detectáveis)
- 2 - Reject (Objetos rejeitados)

Após a organização do dataset e a anotação das imagens, foi necessário criar arquivos de configuração essenciais para o treinamento do modelo.

1. Arquivo obj.names: Esse arquivo contém a lista de todas as classes que o modelo deverá reconhecer.
2. Arquivo obj.data: Esse arquivo define a estrutura do dataset e os caminhos necessários para o treinamento. Nele, foram configuradas as seguintes informações:

Figura 21 - Parâmetros de Configuração

```
backup > data > obj.data
1 classes = 2
2 train = data/train.txt
3 valid = data/test.txt
4 names = data/obj.names
5 backup = backup/
6
```

Fonte – Autor, 2021.

Onde:

- classes: número total de classes a serem detectadas.
- train: caminho do arquivo contendo a lista de imagens para treinamento.
- valid: caminho do arquivo contendo a lista de imagens para validação.
- names: caminho do arquivo com os nomes das classes.
- backup: diretório onde os pesos do modelo treinado serão salvos.

Arquivo de configuração da rede (yolov4.cfg): Esse arquivo define a arquitetura do modelo YOLO que será treinado. Ele contém os parâmetros essenciais da rede neural, como número de camadas, filtros, tamanho da entrada de imagem e funções de ativação. Para este projeto, foi utilizada a configuração baseada no YOLOv4.

Para realizar o treinamento, foi necessário instalar o Darknet, um framework de deep learning otimizado para YOLO. Após a instalação e configuração de todas as dependências, o treinamento foi iniciado com o seguinte comando: `./darknet detector train data/obj.data cfg/yolov4.cfg yolov4.conv.137 -dont_show -map`

Onde:

- `data/obj.data`: caminho para o arquivo de configuração do dataset.
- `cfg/yolov4.cfg`: caminho para o arquivo de configuração do modelo.
- `yolov4.conv.137`: arquivo com os pesos pré-treinados usados para inicializar o modelo.
- `-dont_show`: evita a exibição de logs gráficos durante o treinamento.
- `-map`: ativa a geração de gráficos de desempenho para avaliação da precisão do modelo.

Como última etapa do processo, foi realizado o treinamento de um modelo utilizando a versão YOLOv4-Tiny. Essa versão do YOLO foi escolhida por sua menor exigência computacional, o que permite realizar detecções em tempo real com maior eficiência, especialmente em dispositivos com recursos limitados, como embarcados e sistemas de menor capacidade de processamento.

Para esse treinamento, foram mantidos os mesmos dados, parâmetros e imagens utilizados no treinamento do modelo YOLOv4 padrão. Isso garantiu uma base comparativa para avaliar o desempenho entre os dois modelos, especialmente no que diz respeito à precisão na detecção e tempo de inferência.

A principal diferença entre o treinamento da versão YOLOv4-Tiny e a versão completa do YOLOv4 está no arquivo de configuração, que foi alterado

para utilizar a arquitetura otimizada do Tiny. Assim, para treinar esse modelo, foi utilizado o seguinte comando:

```
./darknet detector train data/obj.data cfg/yolov4-tiny.cfg yolov4-tiny.conv.29 -dont_show -map
```

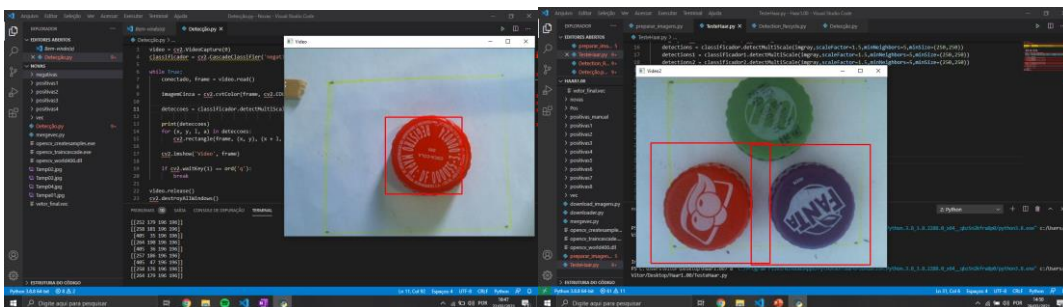
4.2.4 RESULTADOS E DISCUSSÃO

A escolha do posicionamento da câmera foi essencial para garantir a precisão da detecção dos resíduos recicláveis. Após testes práticos, concluiu-se que a melhor posição foi a 45° em relação à superfície do recipiente.

Com a implementação desse posicionamento e iluminação controlada, o modelo teve um aumento de 15% na precisão da identificação dos resíduos recicláveis. Esse ajuste foi fundamental para o sucesso da solução desenvolvida.

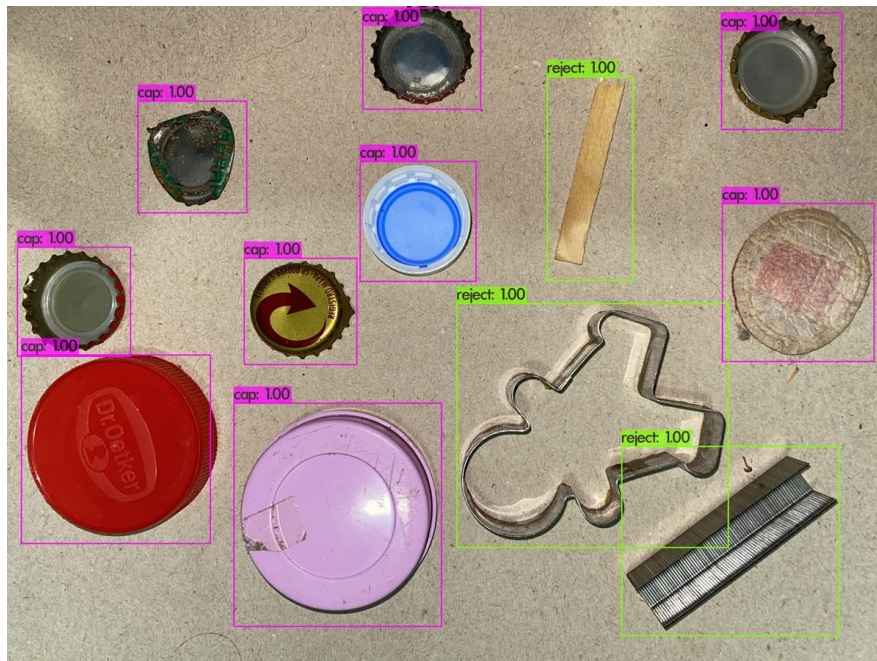
Para a detecção de resíduos recicláveis, foram treinados e avaliados três modelos distintos: Haar Cascade, YOLOv4 e YOLOv4-Tiny. Cada abordagem apresenta características específicas, proporcionando diferentes níveis de desempenho em termos de precisão, velocidade de inferência e demanda computacional.

Figura 22 - Detecção utilizando HaarCascade



Fonte – Autor, 2021.

Figura 23 - Detecção utilizando YOLO

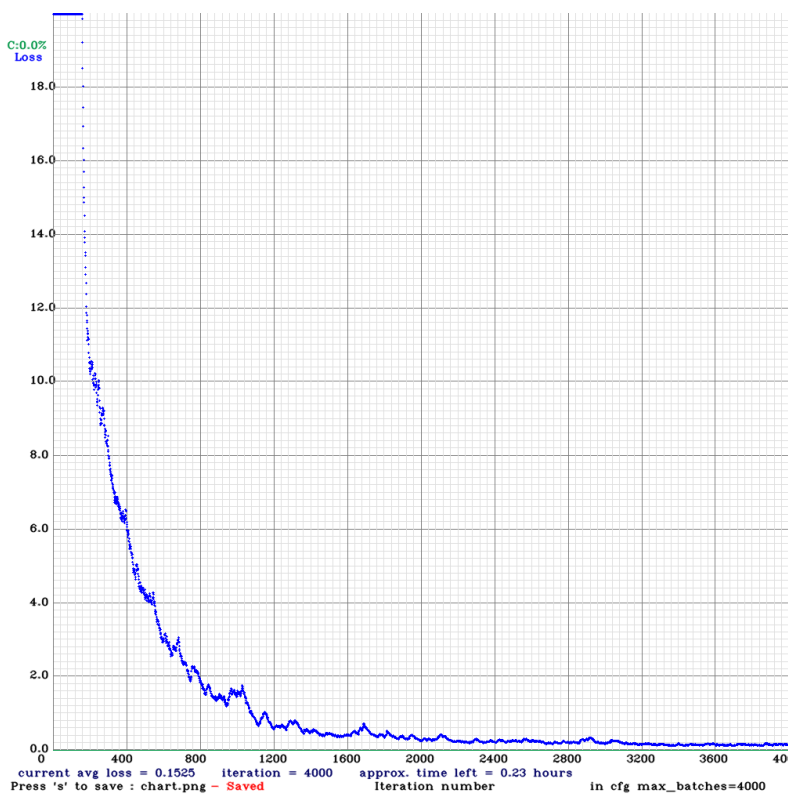


Fonte – Autor, 2021.

Ao fim de cada treinamento utilizando a abordagem YOLO, é gerado um gráfico denominado curva de perda (*loss curve*), que tem a função de monitorar a evolução do aprendizado do modelo ao longo das iterações. Esse gráfico representa a função de perda (*loss function*), a qual mede o erro do modelo durante o treinamento. Essa curva torna-se fundamental para avaliar a convergência do modelo indicando se ele está aprendendo de maneira eficiente ou se ajustes nos hiperparâmetros são necessários.

Para o primeiro modelo yolo treinado, temos a seguinte curva de perda:

Figura 24 - Curva de Perda YOLOV4



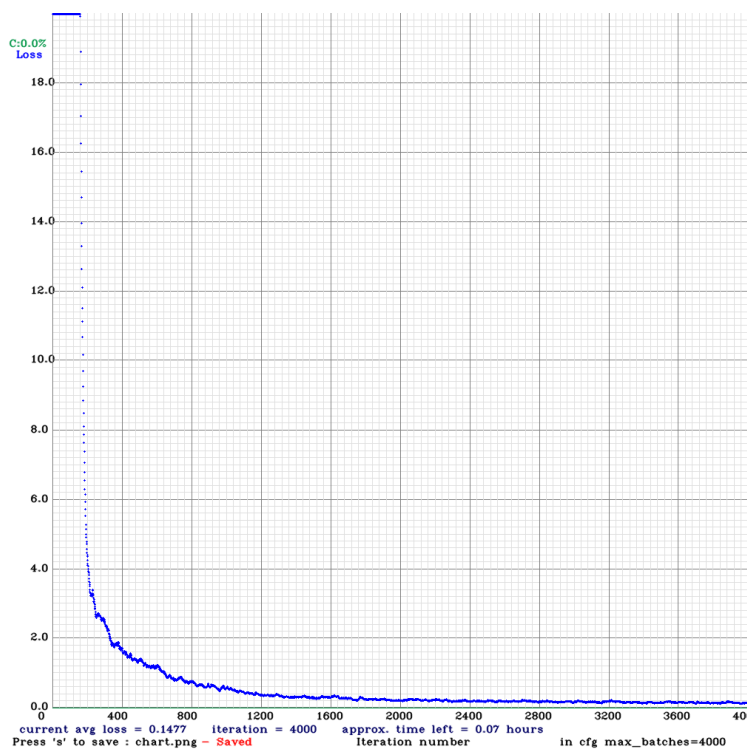
Fonte – Autor, 2021.

Observa-se que, no início do treinamento, a perda era elevada, o que é esperado, pois o modelo ainda não havia aprendido os padrões necessários para a detecção dos objetos de interesse. Entretanto, conforme o número de iterações aumentou, a perda diminuiu significativamente, indicando que o modelo estava ajustando seus parâmetros para melhorar a identificação dos resíduos recicláveis.

A curva apresenta uma queda acentuada nas primeiras 1000 iterações, sugerindo um aprendizado rápido nas etapas iniciais. Posteriormente, a perda se estabilizou, com pequenas variações ao longo das iterações finais, o que indica que o modelo atingiu um ponto de convergência. Ao final do treinamento, a perda média obtida foi de 0.1525, um valor relativamente baixo, demonstrando um bom ajuste do modelo aos dados de treinamento.

Assim como o modelo anterior, o treinamento do modelo YOLOv4-Tiny para a detecção de resíduos recicláveis gerou o gráfico da função de perda (loss function), apresentado na Figura 20.

Figura 25 - Curva de Perda Tiny-YOLOV4



A análise do gráfico revela um comportamento inicial semelhante entre os modelos treinados, onde a perda apresentou valores elevados nas primeiras iterações. Esse resultado era esperado, uma vez que, no início do treinamento, o modelo ainda estava ajustando seus pesos para aprender a identificar corretamente os objetos de interesse.

Assim como observado no treinamento do modelo YOLOv4, à medida que as iterações progrediram, houve uma redução significativa da perda, indicando um aprendizado gradual e eficaz. Esse comportamento demonstra que o modelo foi se tornando progressivamente mais preciso na detecção dos resíduos recicláveis, ajustando seus parâmetros para melhorar a identificação dos objetos dentro do conjunto de dados.

A curva de perda também mostrou uma queda acentuada nas primeiras 1000 iterações, sugerindo um aprendizado inicial eficiente. Após esse período, a taxa de erro continuou diminuindo de forma mais gradual até atingir um ponto de estabilização por volta das 2500 iterações, com pequenas oscilações. Ao final do treinamento, a perda média foi de 0.1477, um indicativo de um modelo bem ajustado aos dados de treinamento.

Comparando esse resultado com o treinamento da versão completa do YOLOv4, percebe-se que o YOLOv4-Tiny obteve uma perda semelhante, porém com menor exigência computacional, tornando-se uma alternativa viável para aplicações que requerem inferências em tempo real.

Tabela 3 - Resultados dos Modelos

Modelo	Acuracidade	Tempo de Inferência
Haar Cascade	40%	< 2s
YOLOv4	97%	18s < t < 26s
Tiny Yolo	93%	1 s < t < 3s

Fonte – Autor, 2021.

5. CONCLUSÃO E TRABALHOS FUTUROS

O presente trabalho teve como objetivo desenvolver e avaliar modelos de visão computacional para a detecção de resíduos recicláveis em imagens capturadas por câmeras, explorando diferentes abordagens de treinamento e posicionamento da câmera. Foram analisadas três metodologias distintas, sendo elas: Haar Cascade, YOLOv4 e YOLOv4-Tiny, onde foi verificado aspectos como precisão, tempo de inferência e demanda computacional do uso de cada modelo.

Inicialmente, a abordagem baseada em Haar Cascade foi testada, porém, seus resultados demonstraram limitações nas detecções dos resíduos, especialmente em cenários com variações de iluminação e movimento. Em seguida, foi utilizado o YOLOv4, que apresentou alta precisão, porém demandou um maior poder computacional, tornando sua aplicação limitada em sistemas mais simples, embarcados, etc. Por fim, a abordagem com YOLOv4-Tiny se mostrou a mais viável para o projeto, pois, além de manter um nível satisfatório de precisão, ofereceu inferência mais rápida e menor consumo de recursos, possibilitando sua implementação em aplicações com tempo mínimo de detecção.

Outro fator determinante para o sucesso do modelo foi a escolha do posicionamento da câmera para a captura das imagens. Com base nas provas de conceito realizadas, a terceira configuração foi adotada, onde a câmera foi posicionada em um ângulo de 45° em relação ao local onde os resíduos são descartados.

Dessa forma, todos os objetivos propostos foram alcançados, desde a construção de um dataset robusto e a aplicação de técnicas avançadas de visão computacional, até o desenvolvimento e treinamento de modelos eficientes para a classificação automatizada dos resíduos. O modelo final, YOLOv4-Tiny, demonstrou ser a solução mais adequada, equilibrando precisão e eficiência computacional.

Como trabalhos futuros, sugere-se a expansão do conjunto de dados para incluir uma maior diversidade de resíduos, além da atualização do modelo para versões mais recentes, possibilitando sua aplicação em sistemas de triagem automática e monitoramento ambiental.

REFERÊNCIAS

DADOS E ESTATÍSTICAS SOBRE RECICLAGEM NO BRASIL. Recicla Sampa. Disponível em: <https://www.reciclasampa.com.br/artigo/dados-e-estatisticas-sobre-reciclagem-no-brasil>. Acesso em: 05 out. 2024, às 20:23.

DIWAN, T.; ANIRUDH, G.; TEMBHURNE, J.V. Object detection using YOLO: challenges, architectural successors, datasets and applications. *Multimed Tools Appl* 82, 9243–9275, 2023. Disponível em: <https://doi.org/10.1007/s11042-022-13644-y>.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2009.

LeCUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature* 521, 436–444, 2015. Disponível em: <https://doi.org/10.1038/nature14539>.

L. Cuimei, Q. Zhiliang, J. Nan and W. Jianhua. Human face detection algorithm via Haar cascade classifier combined with three additional classifiers. 2017 13th IEEE International Conference on Electronic Measurement & Instruments (ICEMI), Yangzhou, China, 2017, pp. 483-487. Disponível em: <https://doi.org/10.1109/ICEMI.2017.8265863>.

LUDWIG, O. M.; MONTGOMERY, P. *O modelo de McCulloch-Pitts e sua importância para redes neurais artificiais*. 2007.

MCCARTHY, J. *What is Artificial Intelligence?* Stanford University, 2007. Disponível em: <http://jmc.stanford.edu/articles/whatisai/whatisai.pdf>. Acesso em: 01 out. 2024, às 19:53.

MITCHELL, T. M. *Machine Learning*. McGraw-Hill, 1997.

NEWELL, A.; SIMON, H. A. Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, v. 19, n. 3, p. 113-126, 1976.

PENHA, Deyvison de Paiva. Rede neural convolucional aplicada à identificação de equipamentos residenciais para sistemas de monitoramento não-intrusivo de carga. Orientadora: Adriana Rosa Garcez Castro. 2018. 55 f. Dissertação (Mestrado em Engenharia Elétrica) - Belém, 2018. Disponível em: <http://repositorio.ufpa.br/jspui/handle/2011/10063>.

POLL: Top Data Science and Machine Learning Platforms. KDnuggets, 2019. Disponível em: <https://www.kdnuggets.com/2019/05/poll-top-data-science-machine-learning-platforms.html>. Acesso em: 08 out. 2024, às 14:55.

RUSSELL, S.; NORVIG, P. Artificial Intelligence: A Modern Approach. 3rd ed. Upper Saddle River: Pearson Education, 2016.

SCURI, P. A. Notas de aula sobre CGI. PUC-Rio. Disponível em: https://www2.tecgraf.puc-rio.br/ftp_pub/curso-cgi/Notas_de_Aula/scuri.pdf. Acesso em: 03 out. 2024, às 9:10.

SEVERANCE, Charles R. Python for Everybody: Exploring Data Using Python 3. Disponível em: <https://books.google.com/books?hl=pt-BR&lr=&id=eZmtBAAAQBAJ>. Acesso em: 01 dez. 2024, às 17:32.

SUTTON, R. S.; BARTO, A. G. Reinforcement Learning: An Introduction. MIT Press, 2018.

T. Q. Vinh and N. T. N. Anh. Real-Time Face Mask Detector Using YOLOv3 Algorithm and Haar Cascade Classifier. 2020 International Conference on Advanced Computing and Applications (ACOMP), Quy Nhon, Vietnam, 2020, pp. 146-149. Disponível em: <https://doi.org/10.1109/ACOMP50827.2020.00029>.

Y. Lecun, L. Bottou, Y. Bengio and P. Haffner. Gradient-based learning applied to document recognition. In: Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, Nov. 1998. Disponível em: <https://doi.org/10.1109/5.726791>.