Universidade Federal do Pará
Instituto de Tecnologia
Programa de Pós-Graduação em Engenharia Elétrica

Estimativa Volumétrica de Resíduos Sólidos Urbanos em Imagem de Visualização Única

Julio Leite Azancort Neto

DM 19/2024

UFPA / ITEC / PPGEE
Campus Universitário do Guamá
Belém - Pará - Brasil
2024

Universidade Federal do Pará

Instituto de Tecnologia

Programa de Pós-Graduação em Engenharia Elétrica

Julio Leite Azancort Neto

Estimativa Volumétrica de Resíduos Sólidos Urbanos em Imagem de Visualização Única

Dissertação de Mestrado submetida à avaliação da Banca Examinadora aprovada pelo colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Pará e julgada adequada para a obtenção do Grau de Mestre em Engenharia Elétrica na área de computação aplicada.

DM 19/2024

UFPA / ITEC / PPGEE

Campus Universitário do Guamá

Belém - Pará - Brasil

2024

A991e Azancort Neto, Julio Leite, 1999-

Estimativa volumétrica de resíduos sólidos urbanos em imagem de visualização única. / Julio Leite Azancort Neto.-2024.

Orientador: Carlos Renato Lisboa Francês.

Dissertação (Mestrado) – Universidade Federal do Pará, Instituto de Tecnologia, Programa de Pós-graduação em Engenharia Elétrica, Belém, 2024.

1. Redes neurais (computação). 2. Processamento de imagens – técnicas digitais. 3. Resíduos sólidos – modelos matemáticos. I. Título.

CDD 23 ed. 006.32

Elaborada por Lucicléa Silva de Oliveira - CRB 2/648

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

“ESTIMATIVA VOLUMÉTRICA DE RESÍDUOS SÓLIDOS URBANOS EM IMAGEM DE VISUALIZAÇÃO ÚNICA”

AUTOR: **JÚLIO LEITE AZANCORT NETO**

DISSERTAÇÃO DE MESTRADO SUBMETIDA À BANCA EXAMINADORA APROVADA PELO COLEGIADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA, SENDO JULGADA ADEQUADA PARA A OBTENÇÃO DO GRAU DE MESTRE EM ENGENHARIA ELÉTRICA NA ÁREA DE COMPUTAÇÃO APLICADA.

APROVADA EM: 02/09/2024

BANCA EXAMINADORA:

Prof. Dr. Carlos Renato Lisboa Francês
(Orientador – PPGEE/ITEC/UFPA)

Prof.ª Dr.ª Jasmine Prsicyla Leite de Araújo
(Avaliadora Interna – PPGEE/ITEC/UFPA)

Prof.ª Dr.ª Evelin Helena Silva Cardoso
(Avaliadora Externa ao Programa – CAMPUS CASTANHAL/UFPA)

VISTO:

Prof. Dr. Diego Lisboa Cardoso
(Coordenador do PPGEE/ITEC/UFPA)

Agradecimentos

Em primeiro lugar, gostaria de expressar minha profunda gratidão à minha família, especialmente aos meus pais, Julimar Azancort e Rozi Azancort, por seu amor incondicional, apoio constante e compreensão ao longo de toda a minha jornada acadêmica e na vida. Sem as oportunidades que me deram e todos os esforços que fizeram, eu nunca teria chegado onde cheguei e nem conquistado o que conquistei.

À minha noiva, Glenda Melo, por estar ao meu lado em todos os momentos, por todas as palavras de apoio e por ter sido forte comigo nos momentos mais difíceis e complexos da minha vida. Sempre serei grato por tudo que fez e espero conseguir retribuir ao menos a metade de tudo que você já fez por mim.

À minha irmã, July Azancort, e minhas sobrinhas Ana Clara Azancort e Aurora Azancort, por sempre torcerem por mim e participarem ativamente da minha trajetória. Espero incentivá-las a sempre lutar para conquistar seus sonhos e vitórias.

À minha querida avó, Erotilde Macedo, que esteve presente em todos os momentos da minha vida e em todas as minhas graduações. Sua presença fez e faz total diferença e me fortalece a buscar realizar tudo aquilo que você desejou mas não teve a oportunidade.

Agradeço imensamente ao meu orientador, Professor Renato Francês, pelas suas orientações, paciência, confiança e valiosas contribuições para a minha jornada acadêmica e pessoal. Sua expertise e conselhos foram fundamentais para o desenvolvimento desta pesquisa e para a realização deste grande sonho.

Gostaria de agradecer aos membros da banca examinadora, Prof.ª Dra. Jasmine Priscyla Leite de Araújo e Prof.ª Dra. Evelin Helena, pelas suas valiosas orientações, críticas construtivas e sugestões enriquecedoras.

Agradeço a todos os meus colegas do Laboratório de Planejamento de Redes de Alto Desempenho (LPRAD), que influenciaram e contribuíram significativamente para o desenvolvimento deste trabalho, em especial ao MSc. Romário Silva por toda a ajuda e contribuições.

Gostaria de agradecer a todos os meus amigos que, de alguma forma, me ajudaram a alcançar este sonho. Não citarei nomes para não esquecer de ninguém, mas saibam que o apoio e companhia de vocês fizeram total diferença na minha vida e nas minhas conquistas.

A todos que, de alguma forma, contribuíram para a realização desta dissertação, o meu muito obrigado.

"To anyone going through tough times: believe me, been there, done that.
But every day above ground is a great day, remember that."

Armando "Pitbull" Pérez

Resumo

A gestão eficiente de resíduos sólidos é crucial para manter a cidade limpa e sustentável. Este trabalho apresenta uma metodologia que utiliza algoritmos bem estabelecidos para a estimativa de volume na gestão de resíduos sólidos urbanos a partir de imagens de única visualização. O sistema proposto baseia-se em conceitos e modelos de visão computacional de última geração, incluindo segmentação de instâncias, estimativa de profundidade e cálculo de volume baseado em nuvem de pontos. A metodologia demonstrou a capacidade de estimar com precisão o volume de objetos de resíduos sólidos, tanto individuais quanto múltiplos, em imagens. A abordagem foi avaliada utilizando dados do mundo real. Apesar dos desafios, como o reescalonamento manual de distâncias e conjuntos de dados limitados, o sistema possui um potencial considerável para refinamento e aprimoramento, visando cenários complexos como os urbanos reais. Os resultados numéricos mostraram que o sistema proposto é promissor mesmo em cenários complexos, com valores de erro percentual absoluto médio (MAPE) de 8,60% para resíduos únicos e 9.23% para resíduos múltiplos, resultando em uma média geral de 8.91%. O coeficiente de determinação foi de 95.11% para instâncias únicas e 87.64% para múltiplas. A metodologia proposta contribui significativamente para o avanço das tecnologias de gestão em *smart cities*.

Palavras-chaves: Visão computacional. Aprendizagem Profunda. Resíduos Sólidos Urbanos. Meio Ambiente. Estimativa de Volume.

Abstract

Efficient solid waste management is crucial for keeping the city clean and sustainable. This work presents a methodology that uses well-established algorithms for volume estimation in urban solid waste management from single-view images. The proposed system is based on state-of-the-art computer vision concepts and models, including instance segmentation, depth estimation, and volume calculation based on point clouds. The methodology demonstrated the ability to accurately estimate the volume of both individual and multiple solid waste objects in images. We evaluated our approach using real-world data. Despite challenges such as manual rescaling of distances and limited datasets, our system shows considerable potential for refinement and improvement, targeting complex scenarios like real urban environments. Numerical results indicated that the proposed system is promising even in complex scenarios, with mean absolute percentage errors (MAPE) of 8.60% for single waste and 9.23% for multiple wastes, resulting in an overall average of 8.91%. The coefficient of determination was 95.11% for single instances and 87.64% for multiple instances. The proposed methodology significantly contributes to the advancement of management technologies in smart cities.

Keywords: Computer Vision. Deep Learning. Urban Solid Waste. Environment. Volume Estimation.

Lista de Figuras

Lista de Tabelas

Lista de abreviaturas e siglas

GRS	Gestão de Resíduos Sólidos
GISR	Gestão Integrada e Sustentável de Resíduos
IA	Inteligência Artificial
IoT	Internet das Coisas
RSU	Resíduos Sólidos Urbanos
CNN	Redes Neurais Convolucionais
MAPE	Erro Percentual Absoluto Médio
RPE	Erro Percentual Relativo
ABREMA	Associação Brasileira de Resíduos e Meio Ambiente
ISLU	Índice de Sustentabilidade da Limpeza Urbana
PNRS	Política Nacional de Resíduos Sólidos
ODS	Objetivos de Desenvolvimento Sustentável
GPS	Sistema de Posicionamento Global
DL	Deep Learning
ML	Machine Learning
SSD	Single Shot Detection
RPN	Region Proposal Networks
YOLO	You Only Look Once
COCO	Common Objects in Context
GPU	Unidade de Processamento Gráfico
ResNet	Residual Network
FPN	Feature Pyramid Network
FC	Fully Connected

MLP	Multilayer Perceptron
ReLU	Unidade Linear Retificada
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
LiDAR	Light Detection and Ranging
SIFT	Scale Invariant Feature Transform
VGG	Visual Geometry Group
MVS	Pyramid Multi-View Stereo
ICP	Iterative Closest Point
RANSAC	RANdom SAmple Consensus
mRVE	Erro Volumétrico Relativo Médio
VIA	VGG Image Annotator
SOR	Statistical Outlier Removal Filter
PCA	Principal Component Analysis
MRCNN	Mask Regional Convolutional Neural Network
R^2	Coeficiente de Determinação

Sumário

1 Introdução

Este capítulo introduz as bases do desenvolvimento do trabalho, abrangendo a contextualização do tema, a descrição do problema, os objetivos gerais e específicos, além de apresentar a estrutura da dissertação.

1.1 Contextualização

A geração de resíduos é uma consequência do aumento da população, da urbanização e do desenvolvimento econômico. Apesar do problema de Gestão de Resíduos Sólidos (GRS), afetar todos os indivíduos e governos em todos os países do mundo, todo esse lixo ainda é mal administrado causando impactos negativos diretamente na saúde pública e no meio ambiente (ARBELáEZ-ESTRADA et al., 2023).

Um dos elementos mais importantes a serem definidos para alcançar uma gestão eficaz de resíduos e proteção ambiental é o estabelecimento de aterros sanitários bem localizados. Como os principais métodos de disposição de resíduos incluem coleta, tratamento, reciclagem e disposição final, é necessário prevenir os perigos e riscos de contaminação a curto prazo (AZANCORT NETO et al., 2021).

Não há dúvidas sobre a importância e o desafio no processo de gestão eficiente de resíduos. O modelo de Gestão Integrada e Sustentável de Resíduos (GISR) apresenta todas as partes necessárias para um processo eficiente e responsável de disposição de resíduos. O modelo é baseado em cinco pontos: Coleta, transferência e transporte; Geração e separação; Tratamento; Reciclagem e Disposição final (GUERRERO; MAAS; HOGLAND, 2013).

No entanto, há três itens muito importantes a serem analisados antes do processo de coleta, que não são mencionados pelo modelo GISR. O processo de identificação da localização, volume e tipo de resíduos a serem coletados. Além disso, estimar com precisão o volume de resíduos é essencial para a coleta, transporte e processamento eficientes (AZANCORT NETO et al., 2024a).

Em seu artigo, Zhang et al. (2022) analisaram o processo de GRS do ponto de vista da separação dos resíduos. O artigo é baseado em quatro tópicos: Geração e separação na fonte; Coleta e transporte; Pré-tratamento e recuperação de recursos. O trabalho também destaca a adoção cada vez mais popular e frequente de Inteligência Artificial (IA), Internet das Coisas (IoT) e tecnologias 5G para resolver problemas relacionados à GRS.

Apesar dos avanços, ainda existe uma lacuna significativa a ser preenchida no processo de coleta de resíduos: a estimação volumétrica. Logo, implementar uma metodologia de estimação volumétrica pode trazer diversas vantagens para otimizar o processo de coleta

e transporte de lixo.

A partir da estimação volumétrica é possível realizar uma alocação mais precisa dos recursos, como veículos de coleta e pessoal, o que pode resultar em uma redução significativa dos custos operacionais. Além disso, possibilita um planejamento mais eficiente das rotas de coleta, evitando sobrecargas e subutilizações dos veículos, o que contribui para a diminuição das emissões de gases poluentes.

Outra vantagem é a melhoria na previsão da demanda de coleta. Com dados mais precisos sobre o volume de resíduos, as empresas de coleta podem ajustar suas operações de acordo com as variações sazonais ou eventuais, garantindo um serviço mais ágil e eficiente. Ademais, a estimação volumétrica pode auxiliar na identificação de áreas com maior geração de resíduos, permitindo a implementação de políticas públicas direcionadas e programas de reciclagem mais eficazes.

No entanto, a estimativa volumétrica de resíduos sólidos urbanos ainda é pouco explorada no meio acadêmico, conforme levantamento realizado até o presente momento. Isso pode ser atribuído à falta de um consenso claro sobre o estado da arte nesse campo. Diversos autores têm explorado o tema utilizando metodologias variadas, predominantemente baseadas em Visão Computacional. Contudo, a complexidade do contexto e a escassez de dados de qualidade e representativos são desafios enfrentados por todos os métodos de estimativa volumétrica, como observado por Lu e Chen (2022), e os resíduos sólidos não escapam dessa realidade.

Essa falta de uniformidade e a presença de obstáculos técnicos e de dados podem ter desencorajado pesquisas mais amplas e aprofundadas nessa área. No entanto, é importante destacar que a resolução desses desafios pode abrir portas para uma gestão mais eficiente e sustentável dos resíduos, trazendo benefícios significativos para o meio ambiente e para a economia.

1.2 Motivação e Desafios

A efetiva gestão de resíduos sólidos enfrenta uma série de desafios, sendo um dos mais proeminentes a necessidade de estimar com precisão o volume desses resíduos (AZANCORT NETO et al., 2024a). A falta de uma metodologia clara e generalista para essa estimativa torna difícil a alocação eficiente de recursos e o planejamento estratégico das operações de coleta e transporte.

A utilização da estimativa volumétrica de Resíduos Sólidos Urbanos (RSU) pode auxiliar significativamente no complexo processo de coleta, transporte e descarte de lixo, comumente utilizado em diferentes países, incluindo o Brasil. Embora a Política Nacional de Resíduos Sólidos exija o fim dos lixões em todo o país, ainda é possível encontrar

lixões em operação. Um exemplo é o aterro localizado no município de Marituba, no Pará, que recebe resíduos sólidos de municípios vizinhos, como da capital Belém, Ananindeua, Benevides e Santa Bárbara. Essas localidades juntas recolhem cerca de 2 mil toneladas de resíduos sólidos por dia (BRITO et al., 2020).

Nesse sentido, para o processo de estimativa, uma das principais dificuldades na criação dessa metodologia reside na heterogeneidade dos resíduos, que podem variar em diversos aspectos como cor, formato, textura, tamanho e diversos outros fatores que variam a composição e agravam o desafio de criar um modelo generalista e eficiente, exigindo assim uma abordagem adaptativa e dinâmica. Outro obstáculo significativo é a escassez de dados representativos e de qualidade, essenciais para validar e aprimorar os métodos de estimativa volumétrica.

Em suma, a estimação volumétrica de resíduos sólidos apresenta desafios significativos devido à sua complexidade e à falta de dados representativos. No entanto, superar esses obstáculos é crucial para uma gestão eficiente e sustentável dos resíduos, com impactos positivos na saúde pública, no meio ambiente e na qualidade de vida das comunidades.

1.3 Objetivos

GERAL

O objetivo deste trabalho é desenvolver um modelo de estimação volumétrica de resíduos sólidos urbanos com base em visão computacional, utilizando Redes Neurais Convolucionais (CNN) para detecção, segmentação e cálculo do volume estimado.

ESPECÍFICOS:

- Realizar o levantamento de técnicas e metodologias utilizadas no processo de estimação de volume através de imagem;
- Identificar as principais lacunas nas soluções propostas;
- Criar um conjunto público de dados (*dataset*) com imagens de resíduos sólidos urbanos, baseado na forma mais comum de descarte de lixo no Brasil, dentro de sacos plásticos. Variando os objetos analisados em suas formas, tamanho, cor, quantidade e complexidade do fundo da imagem;
- Realizar o processo de segmentação semântica em cada um dos objetos utilizados no processo de treino;
- Desenvolver modelos para as tarefas de detecção, segmentação automática, estimação de profundidade e nuvem de pontos, utilizando tecnologias recentes na área de visão computacional;

- Conduzir testes dos modelos para avaliar e validar os resultados alcançados;
- Avaliar o desempenho dos modelos desenvolvidos através das métricas de Erro Percentual Absoluto Médio (MAPE), Erro Percentual Relativo (RPE) e o Coeficiente de determinação.

1.4 Organização da Trabalho

Este trabalho segue a estrutura da seguinte forma:

Capítulo 1: Introduz e contextualiza o problema dos resíduos sólidos urbanos, estabelecendo os objetivos e justificando a relevância da pesquisa proposta.

Capítulo 2: Fornece uma fundamentação teórica abrangente sobre os conceitos e técnicas essenciais utilizados no desenvolvimento desta pesquisa.

Capítulo 3: Apresenta uma revisão das pesquisas correlatas que influenciaram este estudo, destacando suas contribuições significativas e identificando suas limitações.

Capítulo 4: Descreve detalhadamente a metodologia aplicada, incluindo todas as etapas necessárias para alcançar os objetivos estabelecidos nesta dissertação.

Capítulo 5: Exibe os resultados obtidos para os modelos propostos, analisando e discutindo seu desempenho.

Capítulo 6: Oferece as considerações finais do trabalho, sugere direções para futuras pesquisas, lista os trabalhos publicados relacionados ao tema e discute os desafios enfrentados durante a pesquisa.

2 Fundamentação Teórica

2.1 Considerações Iniciais

Neste capítulo, serão apresentados os conceitos mais importantes e necessários para a compreensão da problemática e das tecnologias e técnicas utilizadas no desenvolvimento desta pesquisa. Inicialmente, será realizada uma análise aprofundada do problema relacionado ao descarte incorreto de resíduos sólidos urbanos e a necessidade do envolvimento da tecnologia no processo de gestão desses resíduos. Em seguida, serão discutidos os conceitos fundamentais das técnicas de visão computacional, que desempenham um papel crucial nesta pesquisa. Por fim, serão detalhadas as métricas utilizadas para a avaliação e validação das soluções propostas, garantindo a eficácia e precisão dos métodos desenvolvidos.

2.2 Gestão de Resíduos Sólidos

O crescimento populacional acelerado, altos padrões de vida e altos índices de consumo de bens e energia resultam em elevados níveis de geração de resíduos sólidos urbanos, que representam sérias ameaças ao meio ambiente se não forem descartados ou reciclados de maneira correta e eficaz (NANDA; BERRUTI, 2021a).

Para Nanda e Berruti (2021a), resíduos sólidos urbanos podem ser definidos como lixos e detritos descartados diariamente pela população urbana e rural. Atualmente, estima-se que sejam geradas globalmente 2 bilhões de toneladas de resíduos sólidos urbanos por ano, dos quais cerca de 33% permanecem não coletados pelas prefeituras (Waste Atlas, 2018). Entre os resíduos sólidos urbanos coletados pelas prefeituras, cerca de 70% acabam em aterros sanitários e lixões, 19% são reciclados e 11% são usados para recuperação de energia (NANDA; BERRUTI, 2021a).

De acordo com o Waste Atlas (2018), são geradas em média 0,74 kg de resíduos per capita por dia. O The World Bank (2020) projeta que até 2050 a geração de resíduos sólidos urbanos deve aumentar para 3,4 bilhões de toneladas por dia. Além disso, é previsto que o número de pessoas sem acesso adequado a serviços elementares de gestão de resíduos possa chegar a 5,6 bilhões até 2050 (NANDA; BERRUTI, 2021a).

Os resíduos sólidos urbanos têm origem nos resíduos sólidos de um município, coletados de residências, escritórios, instituições de pequeno porte e empresas comerciais. A composição e classificação dos resíduos sólidos urbanos variam substancialmente entre diferentes países ao redor do mundo e consistem de frações biodegradáveis e não biodegradáveis de materiais orgânicos e inorgânicos, respectivamente. No Brasil, por exemplo, a

forma mais comum de descarte de lixo é feito em simples sacolas plásticas e sem o devido tratamento ou separação para o processo de reciclagem.

Os problemas do não reaproveitamento e descarte incorreto, sobretudo através de sacolas plásticas impróprias para esse fim, é bem explicado por Grechinski (2020). De acordo com o autor, as 8 milhões de toneladas anuais, juntamente com as 80 milhões de toneladas já presentes nos mares, é um problema ambiental e social, sem uma solução única mas que pode ser diminuída através de educação ambiental, na redução do consumo em terra e na criação de soluções para suprir essa demanda na produção de lixo.

Assim como sua composição, as práticas de gestão de resíduos sólidos urbanos diferem entre municípios, cidades, estados e países. No entanto, as etapas básicas na gestão de resíduos sólidos urbanos são: (1) geração; (2) coleta, manuseio e transferência; e (3) disposição, processamento e tratamento (TAşKıN; DEMIR, 2020).

É importante observar que a gestão eficiente dos resíduos sólidos urbanos, ou seja, recuperação ou reciclagem, depende em grande parte da população e do poder econômico do país. Em países desenvolvidos, por exemplo, as tecnologias de conversão de resíduos sólidos urbanos em energia, calor e eletricidade estão bem estabelecidas, de forma a beneficiar o meio ambiente e as pessoas que utilizam esses serviços (MOYA et al., 2017; NANDA; BERRUTI, 2021b).

Por outro lado, nos países em desenvolvimento, onde a densidade populacional exige capacidade disponível para aterros sanitários, ainda há esforços para gerenciar de forma eficaz e higiênica a coleta, transporte e disposição desses RSU. Densidade populacional, aspectos socioeconômicos, facetas culturais, padrões de vida, gestão não planejada e falta de políticas ambientais rigorosas geralmente impedem uma remediação adequada dos resíduos sólidos urbanos nos países em desenvolvimento.

2.2.1 Descarte Irregular de Resíduos

Como comentado anteriormente, o descarte e a coleta inadequada de resíduos urbanos têm diversos impactos negativos na sociedade. A popularização do descarte irregular e o atraso, especialmente por parte de países em desenvolvimento, na utilização de aterros sanitários adequados resultam frequentemente em impactos negativos perceptíveis e ameaças significativas aos recursos hídricos e outros recursos ambientais. Além disso, essa má gestão de resíduos contribui para o aumento dos incômodos ambientais e dos riscos à saúde pública, pois a proliferação de lixões pode facilitar a disseminação de doenças e a contaminação do solo e da água.

Casos como o descrito por Okpara, Kharlamova e Grachev (2021) exemplificam claramente esse tipo de situação. Em seu estudo, os autores descrevem como o descarte e manuseio inadequado de resíduos, juntamente com a natureza costeira topográfica, a

distribuição espacial de lixões e as dinâmicas morfológicas da região que impulsionam o transporte de sedimentos em Port Harcourt, na Nigéria, estão relacionados com a ocorrência de doenças como cólera, diarreia e malária. O estudo demonstrou que a proximidade de lixões irregulares às áreas residenciais aumenta significativamente o risco de surtos dessas doenças, ressaltando a importância de uma gestão eficiente de resíduos e de políticas de saúde pública para mitigar esses riscos. Além disso, o trabalho destaca como as características geográficas específicas da região podem agravar os problemas de saúde pública, sublinhando a necessidade de abordagens contextualizadas para a gestão de resíduos em diferentes localidades.

Na África do Sul os pesquisadores Haywood et al. (2021) analisaram as práticas de descarte de resíduos em comunidades de baixa renda, no período de 2014 a 2019, para entender como as características das habitações influenciam o descarte inadequado de resíduos. De acordo com os autores, famílias que residiam em barracos apresentaram maior propensão a descartar resíduos na rua. Além disso, lares que dependiam de fontes não elétricas para aquecimento ou cozimento, aqueles sem acesso adequado a saneamento e sem água encanada dentro das residências também mostraram maior tendência a descartar resíduos de maneira irregular, seja na rua, no quintal ou por meio de enterramento.

A falta de políticas, interesse e metas para a resolução desse tipo de problemas também é comum no Brasil. De acordo com a Associação Brasileira de Resíduos e Meio Ambiente (ABREMA), que apresentou em 2023 o Índice de Sustentabilidade da Limpeza Urbana (ISLU), o qual pode ser definido como uma ferramenta estatística que tem como objetivo mensurar o grau de aderência dos municípios brasileiro às diretrizes e metas da Lei Federal de Política Nacional de Resíduos Sólidos (PNRS), o índice médio de reciclagem no Brasil não passa dos 3,5% (ABREMA, 2023).

No documento ISLU disponibilizado pela ABREMA (2023), mesmo após 13 anos de PNRS, cerca de 43% das cidades elegíveis continuam destinando o lixo incorretamente, mesmo o prazo de erradicação dos lixões ter expirado em 2014. Além disso, a coleta domiciliar está longe da universalização, deixando de atender cerca de 25% das residências brasileiras

Assim, os resultados do ISLU 2023 demonstram que o Brasil ainda está muito distante de conseguir alcançar as metas dos Objetivos de Desenvolvimento Sustentável (ODS) ligadas a gestão de resíduos sólidos. As estimativas apontam que nenhuma região brasileira alcançara a meta de redução substancial da geração de resíduos.

É nítido a importância da PNRS aos esforços para o tratamento e recuperação (reciclagem) dos RSN, visto que quanto mais alta a taxa de recuperação, melhor é o reaproveitamento dos resíduos coletados. Passando por processos de reciclagem, reutilização e recuperação. Porém, sem o incremento das medidas estruturantes estabelecidas no novo marco legal de saneamento, a tendência é que o Brasil demore muitos anos para atingir as

metas estabelecidas e esperadas no processo de redução do impacto ambiente dos resíduos (ABREMA, 2023)

2.2.2 Impactos Ambientais dos Resíduos Sólidos Urbanos

O impacto da má gestão e falta de incentivos no processo de geração, descarta, coleta e processamento do lixo produzido pode ser sentido em diversas áreas da nossa sociedade, incluindo as pessoas e os ambientes em que vivemos (ESTANQUEIRO JOSé DINIS SILVESTRE; PINHEIRO, 2018). A contaminação do solo, ar e água pode ser facilmente percebida por seus efeitos na flora e fauna, pelo assoreamento dos recursos hídricos e pelo aumento das despesas municipais. Essa situação é preocupante, pois ainda não tem recebido a devida atenção necessária para a sua mitigação (LAWSON et al., 2001; YE et al., 2012; SEROR; HARELI; PORTNOV, 2014).

O objetivo de aprimorar a gestão ambiental e fomentar o desenvolvimento sustentável tem levado os geradores de resíduos a implementarem medidas que diminuam os impactos ambientais (TAM, 2009). Para reduzir esses impactos, é essencial entender suas causas, o que pode ser alcançado por meio de pesquisas sobre a quantidade, composição e fluxo dos resíduos produzidos em uma determinada região (GALLARDO et al., 2015; WU et al., 2016). Essas pesquisas fornecem dados importantes para a tomada de decisões, ajudando a estabelecer tecnologias e ações de gestão e definindo os recursos necessários para serem empregados.

Os sistemas de drenagem frequentemente entupidos por resíduos inadequadamente descartados acentua o problema de inundações durante as estações chuvosas. Quando os canais e bueiros estão obstruídos por resíduos de alimentos, papel, plásticos e outros materiais, a capacidade de escoamento da água é comprometida. Isso não apenas aumenta o risco de inundações em áreas urbanas, mas também pode resultar em danos significativos a propriedades e infraestruturas, além de representar um perigo para a segurança pública (SULEMAN; DARKO; AGYEMANG-DUAH, 2015; JESCA; JUNIOR, 2015; SEPADI, 2022).

Além disso, as condições climáticas adversas, como altas temperaturas e umidade, intensificam a decomposição de resíduos orgânicos descartados, aumentando a produção de líquidos percolados, também conhecido popularmente como chorume. Esses líquidos, compostos por substâncias químicas e materiais em decomposição, infiltram-se no solo e podem contaminar as águas subterrâneas. Esse processo não apenas compromete a qualidade da água, essencial para o abastecimento humano e a vida aquática, mas também pode afetar negativamente os ecossistemas locais, resultando em diversos danos ambientais de longo prazo (UN-HABITAT, 2010; FERRONATO; TORRETTA, 2019).

Como discutido anteriormente, o descarte irregular e não coletado de resíduos

representa uma séria ameaça à saúde pública. A estagnação de água em locais onde são descartados itens como fraldas sujas, alimentos e outros resíduos orgânicos cria um ambiente propício para a proliferação de mosquitos e outros insetos vetores de doenças. Estes insetos podem transmitir doenças graves como dengue, malária, febre amarela, Zika vírus, Chikungunya e outras enfermidades transmitidas por arboviroses, colocando em risco a saúde e bem estar da população local (SULEMAN; DARKO; AGYEMANG-DUAH, 2015; HAYWOOD et al., 2021).

Por fim, o descarte inadequado de resíduos de alimentos, como restos de comida, atrai uma variedade de pragas urbanas, incluindo insetos, roedores e outras pragas que se alimentam desses materiais orgânicos. Essas pragas não apenas representam um incômodo e um problema estético, mas também são portadoras de doenças infecciosas. A presença dessas pragas em áreas urbanas densamente povoadas pode facilitar a propagação de doenças como salmonela, leptospirose e outras infecções gastrointestinais, colocando em risco a saúde pública e exigindo medidas de controle rigorosas (SULEMAN; DARKO; AGYEMANG-DUAH, 2015; HAYWOOD et al., 2021).

2.2.3 Soluções para Coleta de Resíduos

Devido aos motivos e causas discutidos em seções anteriores e aos impactos negativos associados a todos os processos já abordados de geração, manuseio e processamento inadequado de resíduos, há uma necessidade premente de incorporar tecnologias avançadas no desenvolvimento sustentável, considerando o esgotamento dos recursos não-renováveis do planeta. A Inteligência Artificial se destaca como uma modalidade tecnológica fundamental para a gestão socioambiental. A aplicação da IA pode otimizar a utilização de recursos, prever padrões de consumo, melhorar a gestão de resíduos e promover a eficiência energética. Essas capacidades tecnológicas contribuem significativamente para a sustentabilidade ambiental, permitindo um uso mais consciente e eficiente dos recursos disponíveis e reduzindo os impactos negativos no meio ambiente (AZANCORT NETO et al., 2021).

Infelizmente, os departamentos ambientais não têm dedicado a atenção necessária à matematização deste problema, nem investido no desenvolvimento e incorporação de novas tecnologias que visem automatizar e melhorar a eficiência dos processos relacionados aos resíduos gerados em cenários urbanos. Essa falta de enfoque tecnológico e matemático impede a criação de soluções específicas e eficazes para o gerenciamento e disposição desses materiais, exacerbando os impactos ambientais e diminuindo a eficiência das práticas de gestão. A implementação de abordagens mais avançadas e a aplicação de tecnologias emergentes são cruciais para mitigar esses desafios e promover um sistema de manejo mais sustentável e eficiente (PRIYA et al., 2019).

Apesar da falta de investimento e interesse público no desenvolvimento de sistemas e modelos que auxiliem no processo de coleta de resíduos, existem diversos artigos de boa

qualidade disponíveis que têm como objetivo otimizar o processo e reduzir significativamente o custo associado à coleta (SILVA et al., 2023). Isso demonstra que, apesar de complexo e com diversas variáveis, o problema pode ser modelado e solucionado com a utilização de tecnologias como IoT (PARDINI et al., 2019) e algoritmos de *Machine Learning* e *Deep Learning*.

Apesar das práticas de coleta de resíduos variarem de um país para outro, em países em desenvolvimento, a tarefa de coleta é frequentemente realizada de maneira tradicional, com a utilização de caminhões seguindo rotas predeterminadas ou determinadas pelo motorista responsável. Esse cenário apesar de comum pode levar ao uso ineficiente de recursos, isso porque um caminhão pode realizar uma rota com poucos resíduos a serem coletados ou uma com uma quantidade grande demais para ser coletado apenas por aquele veículo. Além do desperdício de tempo e combustível no primeiro cenário, a segunda situação pode levar a problemas ambientais nas cidades, com resíduos sólidos acumulados fora do local apropriado e possíveis problemas como aqueles destacados na seção 2.2.2 (MAHMOOD; ZUBAIRI, 2019).

Neste contexto, novas estratégias para otimizar rotas de coleta de resíduos tem sido exploradas com o objetivo de desenvolver modelos para reduzir o custo de transporte de resíduos (SIDHU et al., 2021; BELLINI et al., 2018). Entre as alternativas utilizadas para essa otimização, as mais populares são aqueles que fazer uso de modelos de previsão para identificar o melhor horário de coleta desses resíduos, juntamente com uso de dados de localização em tempo real para otimizar rotas (THüRER et al., 2016; FERRER; ALBA, 2019).

A primeira estratégia apresentada é menos eficaz devido à natureza estocástica da geração de resíduos, o que dificulta a construção de modelos de previsão com alta precisão. Por outro lado, a utilização de redes físicas de sensores, como redes de sensores sem fio, é a solução mais explorada para otimizar a coleta de resíduos. Em poucas palavras, a estratégia inclui coletar informações sobre o nível de resíduos dentro das lixeiras, enviar essas informações para a nuvem e tomar decisões sobre as rotas com base no nível de resíduos. Esta estratégia é demorada, pois requer o desenvolvimento de um sistema físico para adquirir dados e estudar algoritmos adequados de otimização de rotas. No entanto, geralmente retorna resultados satisfatórios de economia, como mostrado em outros trabalhos (FATANIYA et al., 2019; SOH et al., 2019; APARNA et al., 2021).

Apesar dos avanços na literatura sobre o tema, a estimativa volumétrica no processo de coleta de resíduos sólidos urbanos ainda é um campo pouco explorado. A adoção dessas estimativas pode significativamente otimizar as rotas, evitando cenários de caminhões coletando quantidades insuficientes ou excessivas de resíduos, o que é crucial dado a capacidade limitada desses veículos.

A utilização de modelos de *Deep Learning* e análise de imagem possibilitaria a

análise de diversos tipos de diferentes resíduos de forma automatizada, não dependendo exclusivamente de um único tipo de resíduo para determinar a capacidade de coleta (SILVA et al., 2023). Além disso, a integração desses dados com sistemas de Sistema de Posicionamento Global (GPS) pode facilitar a criação de rotas mais eficientes, permitindo o uso adequado de diferentes tipos e tamanhos de veículos para a coleta. Isso não só reduziria os custos operacionais, mas também minimizaria a pegada de carbono e o impacto ambiental associados à coleta de resíduos.

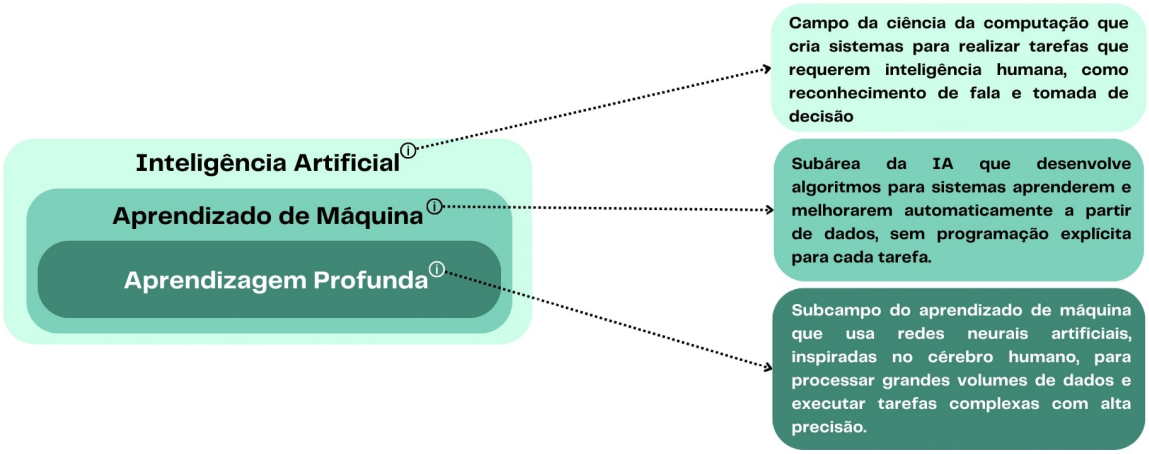
2.3 Aprendizagem Profunda

Aprendizado de máquina, uma subclasse da inteligência artificial, como mostrado na Figura 1. É um aprendizado autônomo baseado em algoritmos, o que significa que o sistema aprende com sua experiência (DONG; WANG; ABBAS, 2021). Por exemplo, o tipo de dado fornecido como entrada ao sistema permite que ele aprenda padrões e responda com base em seu aprendizado na saída. Nesse caso, o sistema se torna mais inteligente com o tempo, sem a intervenção humana. Ele utiliza um algoritmo de aprendizado estatístico que automaticamente aprende e melhora sem ajuda humana, aprendendo através da sua experiência e com um grande banco de dados ou grande quantidade de informações fornecidas na entrada (MATSUO et al., 2022). "Profundo" é o termo que se refere a várias camadas entre a entrada e a saída de uma rede neural, enquanto em redes neurais rasas há, no máximo, duas camadas entre a entrada e a saída da rede neural (DONG; WANG; ABBAS, 2021).

A quantidade significativamente maior de camadas no aprendizado profundo (*Deep Learning*) permite uma melhora substancial na eficiência e no aprendizado geral de representações complexas e abstratas dos dados (JANIESCH; ZSCHECH; HEINRICH, 2021). Essa abordagem avançada de inteligência artificial, baseada em redes neurais artificiais de múltiplas camadas, aprimora a capacidade do modelo de imitar a inteligência humana. Como resultado, as máquinas são capazes de realizar tarefas complexas e reconhecer padrões em diversos tipos de dados, como imagens, textos e muito mais (MATHEW; AMUDHA; SIVAKUMARI, 2021).

A profundidade das camadas permite que o sistema aprenda de maneira autônoma, melhorando continuamente sem a intervenção humana, e utiliza algoritmos de aprendizado estatístico para aprimorar seu desempenho (ROBERTS; YAIDA; HANIN, 2022). A integração dessas técnicas posiciona a IA, o aprendizado de máquina e o aprendizado profundo como disciplinas essenciais para o avanço tecnológico atual.

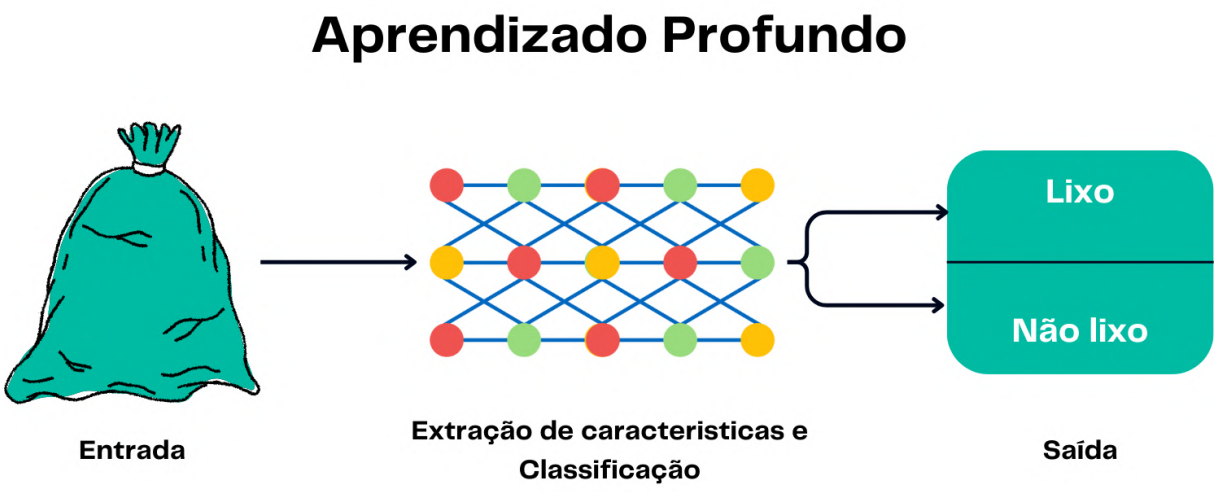
Figura 1 – Conceitos de Inteligência Artificial, Aprendizado de Máquina e Aprendizagem Profunda



Fonte: Elaborado pelo autor

Grande parte do sucesso em IA é atribuído ao aprendizado profundo. O aprendizado profundo utiliza redes neurais artificiais como modelo subjacente para a IA: embora vagamente baseadas em redes neurais biológicas, como o seu cérebro, as redes neurais artificiais são provavelmente melhor compreendidas como uma forma especialmente boa de especificar um conjunto flexível de funções, construídas a partir de muitos blocos computacionais básicos chamados neurônios (ROBERTS; YAIDA; HANIN, 2022). Onde em vez de programar um conjunto específico de instruções para resolver um problema diretamente, os modelos de aprendizado profundo são treinados com dados do mundo real e aprendem como resolver problemas automaticamente, deixando de lado a necessidade de programação manual presente em modelos de profundidade rasa.

Figura 2 – Ilustração de sistema de classificação com aprendizagem profunda



Fonte: Elaborado pelo autor

Atualmente, o aprendizado profundo (DL) é amplamente aplicado em diversas áreas, tornando-se conhecido como um método de aprendizado universal. Ele está sendo empregado em inúmeras situações onde a inteligência artificial pode ser vantajosa, impulsionando avanços tecnológicos significativos, especialmente em cenários onde a intervenção de um especialista humano é limitada (SHARMA; SHARMA; JINDAL, 2021). Entre essas aplicações, destacam-se a visão computacional, o reconhecimento de fala, a compreensão de linguagem, a medicina e muitas outras.

2.3.1 Visão Computacional

A visão computacional, como campo científico interdisciplinar, emprega inteligência artificial para uma compreensão detalhada de dados visuais, seguindo uma abordagem semelhante à dos sistemas visuais humanos. Por meio de modelos e algoritmos, a visão computacional interpreta e analisa dados de imagens ou vídeos, destacando-se no reconhecimento de padrões e identificação de objetos em diversos ambientes (XU et al., 2021). Suas aplicações abordam uma ampla gama de problemas, contribuindo para otimizar ou automatizar tarefas que anteriormente exigiam intervenção manual.

A tecnologia tem sido amplamente utilizada em diversas indústrias, destacando-se os artigos de Esteva et al. (2021), que discute os benefícios, fluxo e desafios da integração em áreas como cardiologia, patologia, dermatologia e oftalmologia. Além disso, o artigo de Kakani et al. (2020) demonstra como a tecnologia está sendo aplicada no processo agrícola e no processamento de alimentos. Esses exemplos ilustram como a visão computacional está sendo adotado em diferentes campos, trazendo avanços significativos e abrindo novas possibilidades em áreas diversas.

A popularização e o avanço tecnológico da visão computacional têm gerado um impacto significativo em várias etapas das indústrias. Seu papel tornou-se crucial, desempenhando uma função essencial na promoção da informatização, digitalização e inteligência dos sistemas de produção industrial. A cada dia, percebemos e somos cada vez mais influenciados pela sua utilização em uma variedade de contextos industriais, o que demonstra a sua importância crescente na melhoria e otimização dos mais diversos processos produtivos (ZHOU; ZHANG; KONZ, 2022).

2.3.1.1 Classificação

O processo de classificação de imagens consiste em categorizar imagens de maneira lógica e metódica em subgrupos, onde cada subgrupo é formado com base nas características comuns presentes nas imagens analisadas Tamuly, Jyotsna e Amudha (2020). Esse processo de indexação é extremamente útil e amplamente utilizado, como, por exemplo, em sites de comércio eletrônico. O aprendizado profundo surge como uma alternativa mais precisa e capaz de automatizar o processo de análise de grandes conjuntos de dados, especialmente

pela sua capacidade de extrair múltiplas características de cada imagem (TAMULY; JYOTSNA; AMUDHA, 2020).

As Redes Neurais Convolucionais, são um dos algoritmos de *Deep Learning* mais populares e amplamente utilizados, especialmente devido à sua alta precisão no processo de classificação de imagens. As CNNs contém duas camadas principais (entrada e saída) e várias outras camadas ocultas. Essas camadas ocultas são compostas por camadas de *pooling*, convolução e normalização e outras camadas totalmente conectadas.

Redes neurais convolucionais têm se mostrado particularmente úteis em tarefas envolvendo o uso de visão computacional, como detecção, classificação e segmentação de objetos. O uso extensivo de CNN é observado atualmente entre pesquisadores médicos para diagnosticar doenças como o Alzheimer e para a classificação de exames de ressonância magnética cerebral (JAIN et al., 2019).

2.3.1.2 Detecção

Assim como na classificação de objetos, a detecção tem sido um tema de pesquisa crucial com os avanços tecnológicos do aprendizado profundo. Com o desenvolvimento de novas ferramentas de aprendizado, agora é mais fácil detectar e analisar a profundidade dos dados (SRIVASTAVA et al., 2021).

No entanto, ao contrário da classificação, onde o objetivo é identificar a categoria de um objeto, na detecção é necessário não apenas identificar a categoria, mas também localizar precisamente a posição do objeto na imagem. Isso pode envolver a detecção de uma ou várias instâncias do mesmo objeto, ou de diferentes objetos. Normalmente, isso é feito utilizando caixas delimitadoras, como ilustrado na Figura 3.

Figura 3 – Exemplos de caixas delimitadoras

Fonte: AZANCORT NETO et al. (2024b)

Com o advento de grandes conjuntos de dados como o COCO *dataset*, houve um aumento na popularidade de algoritmos como SSD, Faster R-CNN e YOLO, que têm

facilitado e aprimorado significativamente o processo de detecção de objetos. O *Single Shot Detection* (SSD) adota uma abordagem "*single shot*", o que significa que realiza a detecção de objetos e a predição de caixas delimitadoras em uma única passagem pela rede neural. Por outro lado, o modelo *Faster Region based Convolutional Neural Networks* (Faster R-CNN) introduziu a ideia de Regiões de Interesse (*Region Proposal Networks* - RPN), que são redes neurais convolucionais dedicadas a gerar propostas de regiões candidatas que possam conter objetos. Por fim, o YOLO (*You Only Look Once)* adota uma abordagem "*end-to-end*", onde a detecção de objetos é realizada em uma única passagem pela rede neural convolucional. Ele divide a imagem em uma grade de células e prevê caixas delimitadoras e probabilidades de classes diretamente para cada célula, simplificando o processo de detecção e tornando-o extremamente rápido (SRIVASTAVA et al., 2021).

2.3.1.3 Segmentação

A segmentação de imagens tem sido fundamental desde o início do desenvolvimento de técnicas relacionadas a visão computacional. Seu funcionamento se baseia no processo de partição ou subdivisões de imagens (ou quadros de vídeo) em diferentes segmentos e objetos, desempenhando um papel central em diversas aplicações como análise de imagens médicas, veículos autônomos, realidade aumentada e outros (MINAEE et al., 2021). As subdivisões geralmente utilizam uma característica visual da cena observada, podem ser cor, textura ou contornos.

Conforme destacado por Graikos et al. (2020), que empregou um sistema baseado em modelo em seu processo de estimação de volume, observa-se que os métodos de segmentação são pouco discutidos ou não são devidamente explicados pelos autores, apesar de não explicar devidamente a causa dessa falta de detalhes.

Segundo Yang et al. (2021), apesar da ampla variedade de algoritmos bem estabelecidos para segmentação, não é comum detalhar os métodos empregados em estudos. Geralmente, os autores utilizam técnicas diretas já populares ou consolidadas para resolver os problemas em questão, ou omitem completamente os detalhes sobre os métodos de segmentação utilizados.

A segmentação na visão computacional desempenha um papel crucial na identificação de objetos e na delimitação de seus contornos em uma imagem, permitindo uma análise e compreensão precisas da cena. Existem duas técnicas fundamentais: a segmentação semântica e a segmentação de instâncias. A segmentação semântica classifica cada pixel em uma categoria, identificando regiões de objetos da mesma classe, mas sem distinguir instâncias individuais. Por outro lado, a segmentação de instâncias identifica e delimita cada instância individual de um objeto, mesmo que pertença à mesma classe, diferenciando, por exemplo, cada pedestre em uma imagem de rua (HAFIZ; BHAT, 2020).

De acordo com (HAFIZ; BHAT, 2020), a segmentação semântica tem como objetivo

obter uma compreensão detalhada da imagem atribuindo um rótulo específico a cada pixel, sendo que cada pixel é designado com um rótulo de classe com base no objeto ou região a que pertence. Enquanto isso, a segmentação de instâncias atribui rótulos distintos a instâncias individuais de objetos que pertencem à mesma classe de objeto. Em outras palavras, a segmentação de instâncias pode distinguir entre diferentes instâncias da mesma classe.

Com o surgimento do aprendizado profundo (LECUN; BENGIO; HINTON, 2015; GOODFELLOW; BENGIO; COURVILLE, 2016), especialmente das CNNs (KRIZHEVSKY; SUTSKEVER; HINTON, 2017; HUANG et al., 2017; LECUN et al., 1998), foram propostas diferentes e únicas metodologias de segmentação de instâncias, como as de Liu et al. (2017), He et al. (2017) e Li et al. (2017).

A popularização de conjuntos de dados de larga escala, como o *Common Objects in Context* da Microsoft, conhecido como COCO (LIN et al., 2014), que contêm cerca de 330 mil imagens de diversas variedades, facilitou o aprimoramento técnico de algoritmos de segmentação de instâncias, melhorando suas performances e eficiência. Exemplos desses algoritmos incluem Mask R-CNN (HE et al., 2017), Fast/Faster R-CNN (GIRSHICK, 2015; REN et al., 2016) e a *Feature Pyramid Network* (FPN) (LIN et al., 2017).

Su et al. (2020) criaram uma representação volumétrica empregando múltiplas camadas de representação. A combinação do volume de custo captura informações em 3D, e a segmentação 2D do objeto foi utilizada para alcançar uma síntese abrangente de objetos complexos.

Conforme o avanço na compreensão e aplicação da segmentação de imagens na visão computacional, é evidente que seu papel fundamental no processo de identificação e delimitação de objetos em uma cena. Ela serve como a espinha dorsal para uma variedade de aplicações vitais da indústria e cada vez mais popular no dia a dia das pessoas. A contínua inovação e pesquisa nessa área prometem não apenas aprimorar a precisão e eficiência dos algoritmos existentes, mas também abrir novas fronteiras de exploração, integrando-se cada vez mais com outras disciplinas. Como resultado, a segmentação de imagens não é apenas uma ferramenta técnica, mas um pilar crucial para desbloquear o potencial total da inteligência artificial e da computação visual.

2.4 Redes Neurais Convolucionais

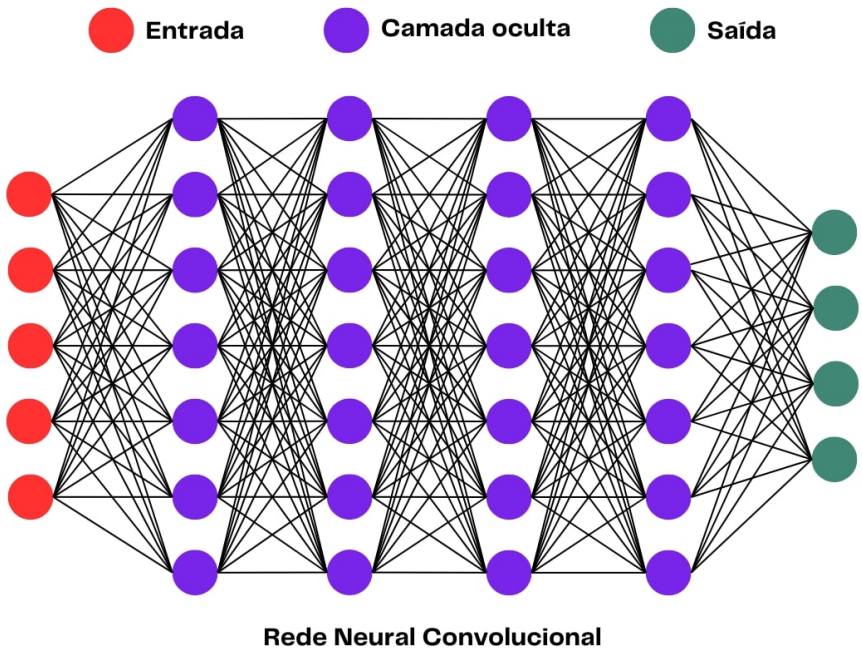
As CNNs têm alcançado feitos notáveis como uma das redes mais representativas do campo de aprendizado profundo, ideais para tarefas baseadas em visão computacional. Sua estrutura de camadas possibilitou realizações anteriormente consideradas impossíveis no campo de visão computacional, como o reconhecimento facial, veículos autônomos, supermercados de autoatendimento e tratamentos médicos inteligentes (KETKAR; MOOLAYIL,

2021).

As redes neurais convolucionais são um tipo de rede neural *feedforward* que extrai características dos dados utilizando estruturas de convolução. Diferentemente dos métodos tradicionais de extração de características, as CNNs não requerem extração manual de características (LI et al., 2022).

Sua arquitetura é inspirada na percepção visual, onde um neurônio biológico corresponde a um neurônio artificial, os núcleos da CNN representam diferentes receptores que podem responder a várias características, e as funções de ativação simulam a transmissão de sinais elétricos neurais, que só são transmitidos ao próximo neurônio se excederem um determinado limiar(KETKAR; MOOLAYIL, 2021). Funções de perda e otimizadores são mecanismos criados para ensinar o sistema CNN a aprender conforme esperado. A Figura 4 demonstra visualmente a estrutura do fluxo de processamento de uma CNN.

Figura 4 – Estrutura de uma Rede Neural de Aprendizado Profundo



Fonte: Elaborado pelo autor

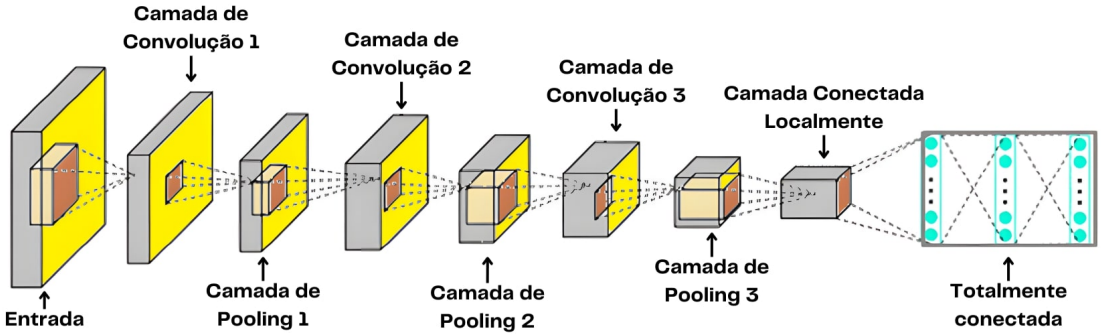
2.4.1 Arquitetura de uma CNN

A arquitetura de uma Rede Neural Convolucional é projetada para processar dados com estrutura de grade, como imagens, preservando as relações espaciais entre os *pixels*. Cada neurônio em uma CNN recebe entradas de um conjunto restrito de neurônios da camada anterior, conhecido como campo receptivo, e aplica uma operação de convolução com um filtro (ou *kernel*) para produzir um mapa de características. Esse processo permite que a CNN detecte padrões locais, como bordas e texturas, nas camadas iniciais e, à

medida que a informação passa por camadas subsequentes, essas características básicas são combinadas para formar representações mais abstratas e complexas, como formas e objetos completos (KETKAR; MOOLAYIL, 2021).

Os neurônios também aplicam funções de ativação que introduzem não-linearidades, permitindo que a rede aprenda representações diversificadas e complexas. A hierarquia de características é uma característica fundamental das CNNs, com camadas iniciais capturando características de baixo nível e camadas intermediárias e finais capturando características de alto nível. Essa hierarquia permite que a CNN construa uma compreensão detalhada da entrada, essencial para tarefas como classificação, detecção e segmentação de imagens (LI et al., 2022). A Figura 5 demonstra uma representação da arquitetura de uma CNN com múltiplas camadas.

Figura 5 – Representação da arquitetura de uma CNN com múltiplas camadas



Fonte: Adaptado de Dialogo (2021)

A eficiência das CNNs é amplificada pelo compartilhamento de pesos, onde um conjunto de filtros é aplicado em todas as posições da entrada, reduzindo o número de parâmetros que precisam ser aprendidos e permitindo uma generalização mais eficaz para novos dados. Em resumo, a estrutura das CNNs explora relações espaciais nos dados de entrada e constrói uma hierarquia de características para uma compreensão profunda das imagens, utilizando operações de convolução, funções de ativação e compartilhamento de pesos para alcançar alto desempenho em tarefas de visão computacional (KETKAR; MOOLAYIL, 2021).

2.4.1.1 Camada Convolucional

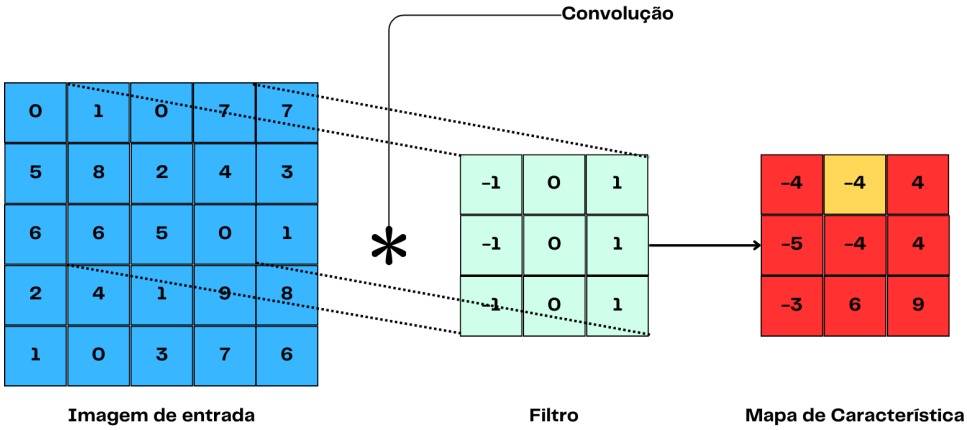
A camada convolucional é o componente central das Redes Neurais Convolucionais e responsável pela maior parte da extração de características dos dados de entrada. Cada camada convolucional consiste em vários filtros (ou *kernels*) que se deslocam pela entrada, aplicando uma operação de convolução para produzir mapas de características (DHILLON; VERMA, 2019). Esses filtros são matrizes de pesos que são aprendidos durante o treinamento da rede e são projetados para detectar padrões locais específicos,

como bordas, texturas, e outros elementos visuais simples. Cada filtro gera um mapa de características distinto, que realça diferentes aspectos da entrada.

O processo de convolução é fundamental para a eficiência das CNNs. Ao restringir a operação a um pequeno campo receptivo da entrada, a camada convolucional consegue capturar detalhes locais com alta precisão. À medida que múltiplas camadas convolucionais são empilhadas, a rede é capaz de combinar essas características locais para formar representações mais complexas e abstratas. Por exemplo, enquanto as camadas iniciais podem detectar bordas e texturas, as camadas mais profundas podem reconhecer formas específicas e objetos inteiros. Essa hierarquia de características permite que a CNN compreenda de forma detalhada e estruturada o conteúdo da imagem (RAITOHARJU, 2022).

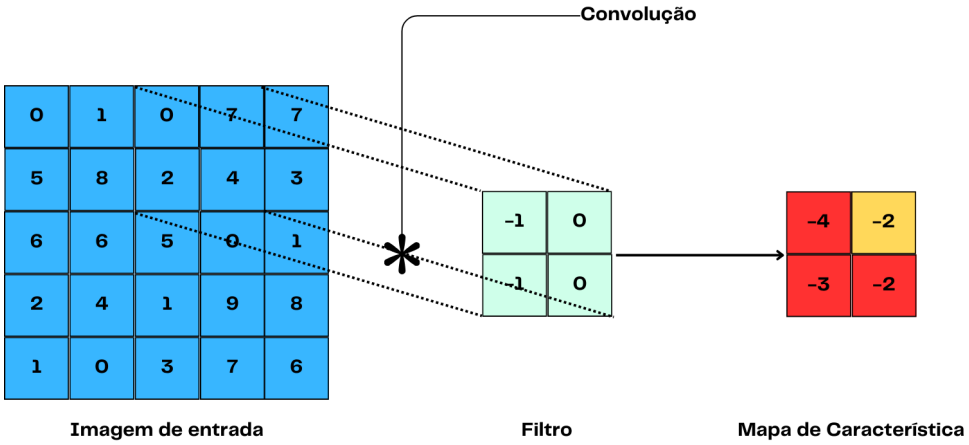
Além da operação de convolução, as camadas convolucionais também utilizam técnicas como *padding* e *stride* para controlar a dimensão dos mapas de características resultantes. *Padding* adiciona pixels adicionais nas bordas da entrada para preservar as dimensões durante a convolução, enquanto *stride* controla o passo do filtro sobre a entrada, afetando a resolução espacial do mapa de características. Essas técnicas, combinadas com a capacidade dos filtros de compartilhar pesos, tornam as camadas convolucionais extremamente poderosas e eficientes para a análise de imagens, permitindo que as CNNs sejam aplicadas com sucesso em uma ampla variedade de tarefas de visão computacional (KETKAR; MOOLAYIL, 2021). As figuras, 6, 7 e 8, demonstram a operação de uma convolução normal, com *stride* e com *padding*, respectivamente.

Figura 6 – Exemplo de operação de convolução



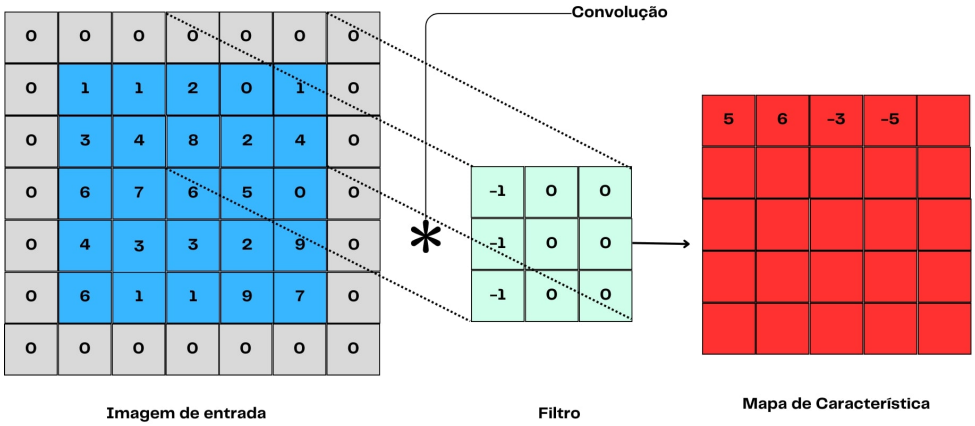
Fonte: Acervo do autor 2024

Figura 7 – Exemplo de operação de convolução com *stride*=(2,3)



Fonte: Acervo do autor 2024

Figura 8 – Exemplo de operação de convolução com *padding*



Fonte: Acervo do autor 2024

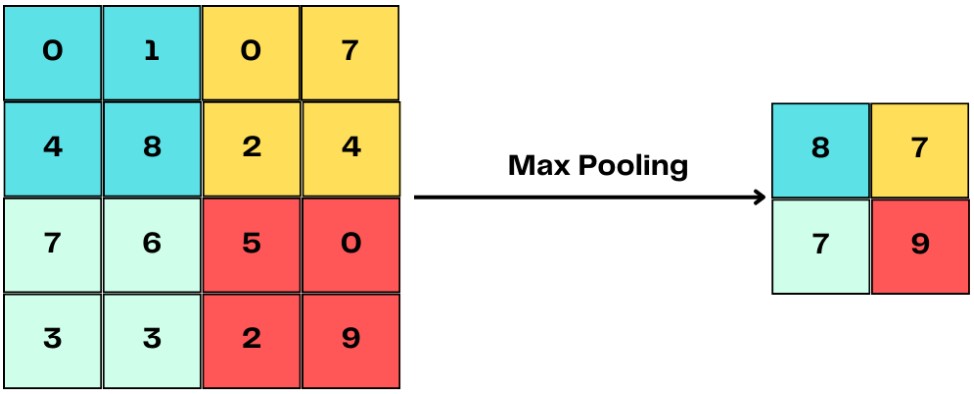
2.4.1.2 Camada de Pooling

A camada de *pooling* é uma componente crucial nas Redes Neurais Convolucionais e desempenha um papel importante na redução da dimensionalidade dos mapas de características, mantendo as informações essenciais. A operação de *pooling* é aplicada após uma ou mais camadas convolucionais e tem como objetivo diminuir o número de parâmetros e o custo computacional da rede, além de ajudar a prevenir o *overfitting*. Existem diferentes tipos de *pooling*, sendo a *max-pooling* a mais comum. Na *max-pooling*, o valor máximo de uma pequena região do mapa de características é selecionado, destacando as características mais salientes (KHAN et al., 2020; GHOSH et al., 2020).

A redução da dimensionalidade através da camada de *pooling* não só torna o modelo mais eficiente, mas também torna a rede mais robusta a variações e distorções nas entradas, como translações e rotações. Isso acontece porque a *pooling* agrega informações locais em uma forma compacta, preservando a presença de uma característica, independentemente

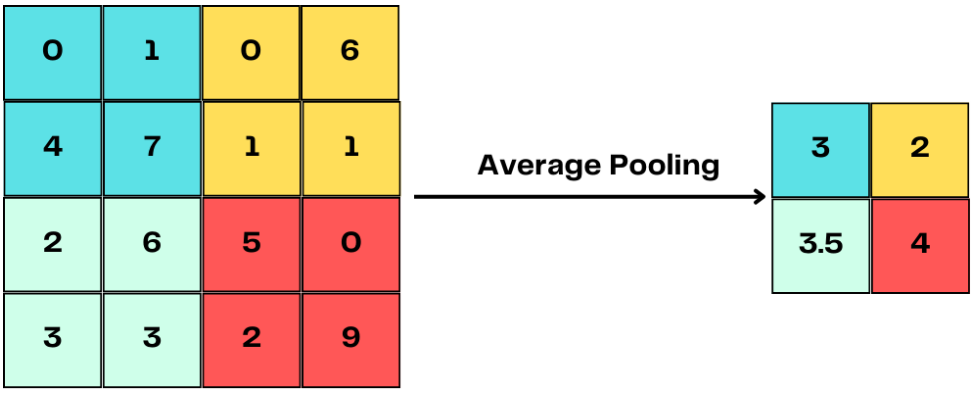
de sua localização exata. Além da *max-pooling*, outras variações como *average pooling*, que calcula a média dos valores em uma região específica, também são usadas, dependendo do contexto e da necessidade da aplicação (GHOSH et al., 2020; KETKAR; MOOLAYIL, 2021). As figuras 9 e 10 demostram exemplos do funcionamento dessas variações.

Figura 9 – Exemplo de *max-pooling*



Fonte: Acervo do autor 2024

Figura 10 – Exemplo de *average pooling*



Fonte: Acervo do autor 2024

Em resumo, a camada de *pooling* contribui significativamente para a eficiência e a robustez das CNNs. Ao reduzir a resolução dos mapas de características e, ao mesmo tempo, manter as informações essenciais, ela permite que as camadas subsequentes processem dados mais compactos e relevantes. Essa compactação facilita a aprendizagem de características mais complexas em camadas posteriores e melhora a capacidade da rede de generalizar bem para novos dados, tornando as CNNs mais eficazes em diversas tarefas de visão computacional.

2.4.1.3 Camada Totalmente Conectada

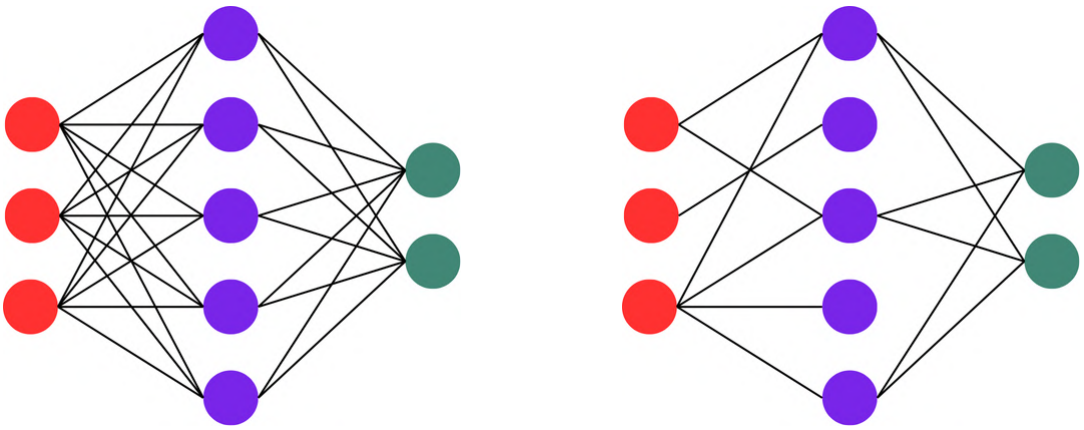
As camadas totalmente conectadas, ou *fully connected* (FC) *layers*, são componentes essenciais nas Redes Neurais Convolucionais, especialmente nas etapas finais da rede.

Nessas camadas, cada neurônio está conectado a todos os neurônios da camada anterior, de maneira semelhante ao que ocorre nas redes neurais tradicionais. Essa conexão densa permite que a rede integre todas as características extraídas nas camadas convolucionais e de *pooling*, sintetizando-as em uma representação final utilizada para realizar tarefas como classificação ou regressão (GHOSH et al., 2020; ZHAO et al., 2024).

A camada totalmente conectada é uma operação global, ao contrário das camadas de convolução e *pooling*, e é tipicamente empregada no final da rede para classificação. Semelhante a uma Rede Neural Perceptron Multicamadas (MLP) (ISABONA et al., 2022), cada neurônio na camada totalmente conectada está conectado a todos os neurônios nas camadas anteriores. Após várias operações de convolução e *pooling*, o mapeamento de características obtido é suficiente para reconhecer as características da imagem. O próximo passo é realizar a classificação, que é a função principal da camada totalmente conectada (GHOSH et al., 2020).

Figura 11 – Diferença entre uma camada totalmente conectada e uma camada de desativação

(a) Camada totalmente conectada (b) Camada de desativação



Fonte: Acervo do autor

Geralmente, a CNN reúne os múltiplos mapeamentos de características obtidos no final em um vetor longo e o envia para a camada totalmente conectada, seguida pela camada de saída para a classificação. A camada totalmente conectada pode integrar informações locais distintivas de classe, capturadas nas camadas de convolução e *pooling* (SAINATH et al., 2013). Embora as camadas totalmente conectadas adicionem uma quantidade significativa de parâmetros à rede, elas são cruciais para a eficácia das CNNs, permitindo uma combinação complexa e não-linear das características aprendidas, resultando em uma maior capacidade de representar padrões complexos nos dados (GHOSH et al., 2020).

2.4.2 Exemplos de Redes Neurais Convolucionais

Como mencionado anteriormente, os algoritmos e modelos baseados em Redes Neurais Convolucionais são altamente eficazes e amplamente utilizados na área de visão computacional. Esses avanços tecnológicos têm levado ao desenvolvimento de uma série de algoritmos inovadores, cada um com suas próprias características e objetivos específicos. Entre os mais notáveis estão LeNet-5, AlexNet, VGGNet, GoogLeNet e ResNet. Cada um desses modelos foi criado para superar desafios específicos e introduzir melhorias em relação aos seus predecessores.

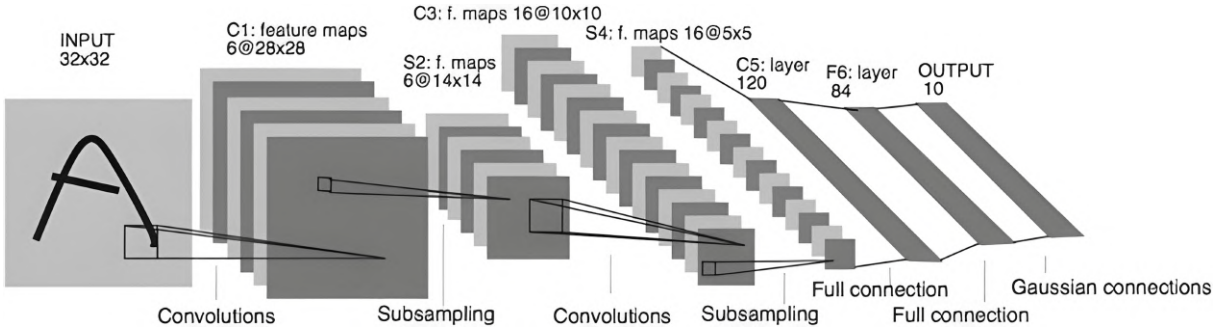
A crescente demanda por algoritmos mais precisos e eficientes na análise e reconhecimento de imagens tem impulsionado a contínua evolução das CNNs. Esses modelos se tornaram essenciais em uma variedade de aplicações, desde o reconhecimento facial e a análise de imagens médicas até a condução autônoma e a segurança pública. A eficácia das CNNs em lidar com grandes volumes de dados e extrair características complexas das imagens as torna uma ferramenta indispensável na era da inteligência artificial.

2.4.2.1 LeNet-5

O LeNet-5 é uma das primeiras arquiteturas de Redes Neurais Convolucionais, projetada para classificar dígitos manuscritos, introduzida por LeCun et al. (1998). A arquitetura do LeNet-5 possui 5 camadas ponderadas (treináveis), sendo três camadas convolucionais e duas camadas totalmente conectadas (FC). As duas primeiras camadas convolucionais são seguidas por camadas de max-pooling para subamostrar os mapas de características, e a última camada convolucional é seguida por duas camadas totalmente conectadas. A última dessas camadas é usada como o classificador, capaz de categorizar 10 dígitos. Esta estrutura foi aplicada com sucesso ao conjunto de dados MNIST, um banco de dados padrão de dígitos manuscritos (LECUN et al., 1998; GHOSH et al., 2020).

O LeNet-5 foi um marco significativo no campo da aprendizagem profunda e visão computacional, demonstrando a capacidade das CNNs de aprender e generalizar padrões complexos a partir de dados de entrada. Sua arquitetura simples e eficaz foi amplamente utilizada como base para o desenvolvimento de modelos mais complexos e robustos. A simplicidade do LeNet-5 permitiu uma redução significativa na dimensionalidade dos dados, preservando as características essenciais para uma classificação precisa, e influenciou o desenvolvimento de arquiteturas modernas em uma ampla variedade de aplicações, desde o reconhecimento de objetos até a análise de imagens médicas e a condução autônoma.

Figura 12 – Arquitetura da CNN LeNet-5



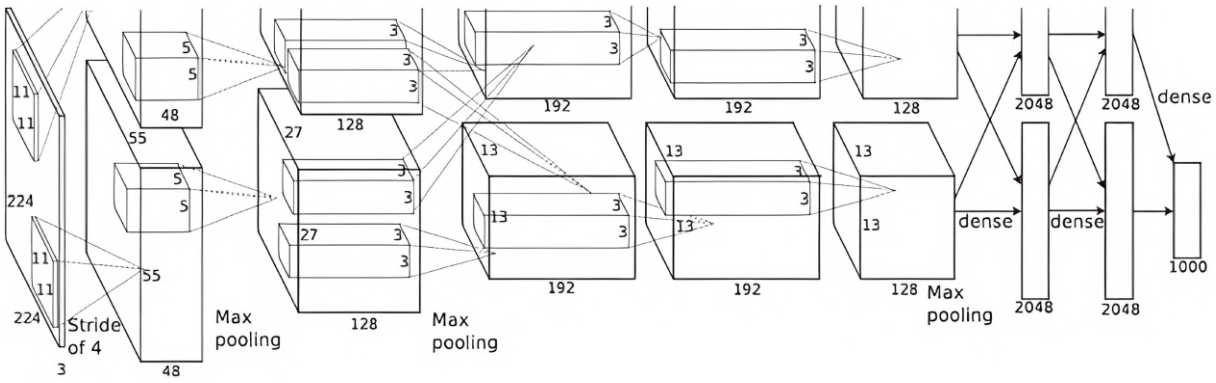
Fonte: LeCun et al. (1998)

2.4.2.2 AlexNet

Inspirado pelo LeNet, Krizhevsky et al. desenvolveram o primeiro modelo de CNN em grande escala, chamado AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012). Projetado para classificar dados do ImageNet, o AlexNet consiste em oito camadas ponderadas (aprendíveis), das quais as primeiras cinco são camadas convolucionais, seguidas por três camadas totalmente conectadas. A camada de saída final é projetada para classificar as imagens de entrada em uma das mil classes do conjunto de dados ImageNet, utilizando 1.000 unidades (KRIZHEVSKY; SUTSKEVER; HINTON, 2012; GU et al., 2018).

O AlexNet representou um avanço significativo no campo da visão computacional e do aprendizado profundo, demonstrando a capacidade das CNNs de lidar com grandes volumes de dados e tarefas complexas de classificação. Ele introduziu várias inovações, como o uso de ReLU (Unidade Linear Retificada) para ativação, que acelerou o treinamento da rede, e o *dropout* para prevenir o *overfitting*. Além disso, o AlexNet explorou a paralelização em GPUs, permitindo a manipulação eficiente de grandes conjuntos de dados. Este modelo foi um ponto de virada na competição *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC) em 2012, onde superou significativamente os métodos anteriores, estabelecendo novos padrões para a classificação de imagens e influenciando o desenvolvimento de futuras arquiteturas de CNN (KHAN et al., 2020).

Figura 13 – Arquitetura da CNN AlexNet



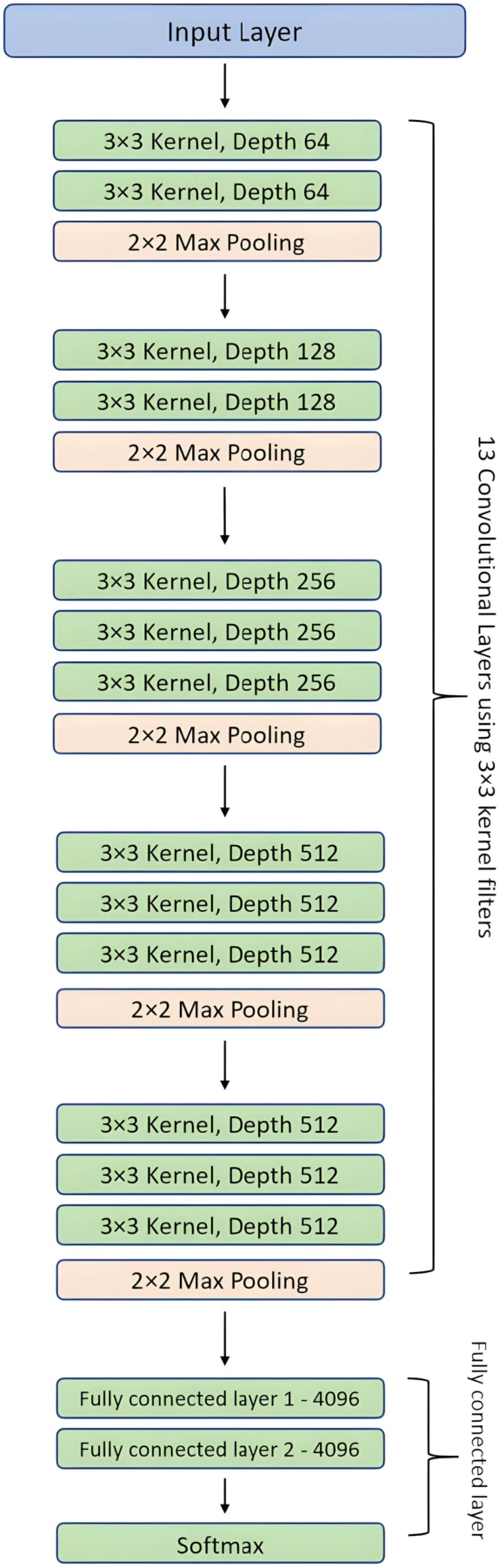
Fonte: Krizhevsky, Sutskever e Hinton (2012)

2.4.2.3 VGGNet

O VGGNet, uma das arquiteturas de CNN mais influentes, foi introduzido por Simonyan e Zisserman (2014). Os autores apresentaram seis configurações diferentes de CNN, sendo as mais bem-sucedidas o VGGNet-16 (configuração D) e o VGGNet-19 (configuração E). O VGGNet é notável por sua simplicidade e profundidade, utilizando filtros de convolução pequenos (3x3) ao longo de toda a rede. Esta abordagem permite que a rede tenha um grande número de camadas de peso, contribuindo para seu excelente desempenho em várias tarefas de reconhecimento de imagens (SIMONYAN; ZISSERMAN, 2014; GHOSH et al., 2020).

A arquitetura do VGGNet consiste em uma série de camadas convolucionais seguidas por camadas de max-pooling e, finalmente, camadas totalmente conectadas para a classificação. O VGGNet-16 e o VGGNet-19 referem-se ao número de camadas de peso (16 e 19, respectivamente) na rede. Esses modelos têm sido amplamente utilizados e adaptados devido à sua eficácia na captura de características de imagens e à sua relativa simplicidade em comparação com arquiteturas mais recentes e complexas.

Figura 14 – Arquitetura da CNN VGGNet



Fonte: Tammina (2019)

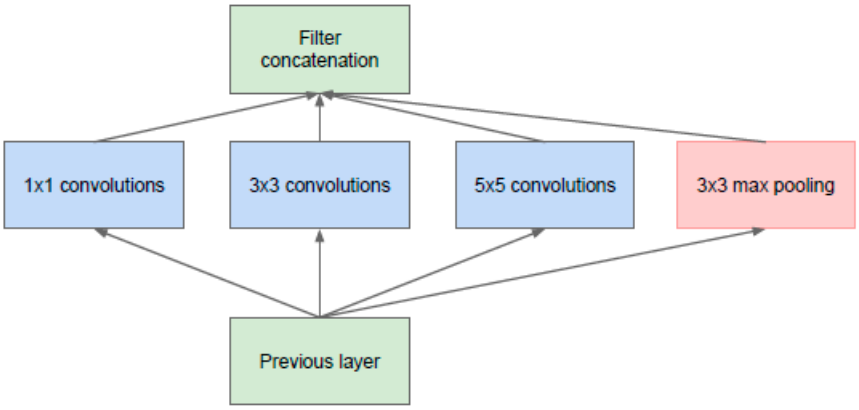
2.4.2.4 GoogLeNet

O GoogleNet, ou Inception v1, é uma arquitetura revolucionária de rede neural convolucional desenvolvida por Szegedy et al. (2015), notável por sua abordagem inovadora em relação aos modelos convencionais. Em vez de seguir uma arquitetura sequencial linear, o GoogleNet utiliza módulos de Inception como blocos fundamentais. Cada módulo de Inception opera em paralelo, combinando saídas de camadas de convolução de diferentes tamanhos de filtro (1x1, 3x3 e 5x5). Isso resulta em características de saída de alta dimensionalidade que podem ser complexas de processar. Para mitigar isso, o GoogleNet incorpora técnicas de redução de dimensionalidade dentro dos módulos de Inception, como mostrado na figura 15 (b), otimizando o fluxo de informação e melhorando a eficiência computacional em comparação com a versão inicial menos refinada, ilustrada na figura 15 (a) Szegedy et al. (2015), Ghosh et al. (2020).

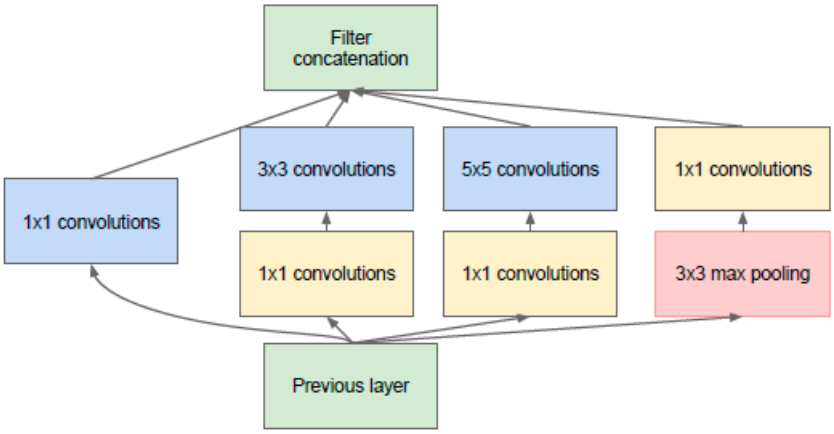
A estrutura do GoogleNet compreende 22 camadas ponderadas, com múltiplos módulos de Inception empilhados sequencialmente. Essa abordagem não apenas permite uma exploração mais profunda e eficaz das características das imagens, mas também ajuda a mitigar problemas de *overfitting*, comuns em redes profundas. Ao combinar eficientemente diferentes operações de convolução em paralelo, o GoogleNet demonstrou um desempenho significativamente melhor na classificação de imagens em comparação com outras arquiteturas contemporâneas da época, marcando um avanço fundamental no desenvolvimento de CNNs para visão computacional.

Figura 15 – Arquitetura da CNN GoogLeNet

(a) Inception Simples



(b) Módulo Inception com redução de dimensionalidade



Fonte: Szegedy et al. (2015)

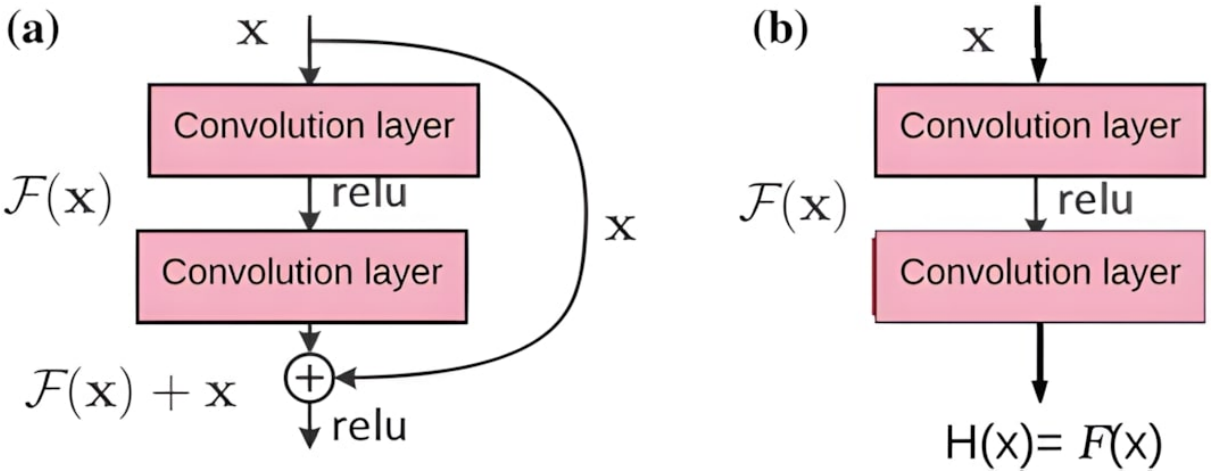
2.4.2.5 ResNet

O ResNet, ou *Residual Network*, foi proposto por He et al. (2016) da Microsoft em resposta aos problemas de gradientes desvanecentes em modelos de CNN profundos. O ResNet introduziu o conceito de conexões de salto de identidade para mitigar esses problemas. Em vez de aprender um mapeamento direto $H(x) = F(x)$, onde $F(x)$ é a transformação desejada, o ResNet adota um mapeamento residual $H(x) = F(x) + x$. Isso permite que a rede aprenda as diferenças residuais entre as ativações esperadas e as ativações atuais, facilitando o treinamento de redes extremamente profundas.

A arquitetura do ResNet é composta por vários blocos residuais, cada um contendo camadas de convolução 3x3. A Figura 16 ilustra a diferença entre o mapeamento direto (a) e o mapeamento residual (b). Esses blocos residuais não só facilitam o fluxo de gradientes através da rede, combatendo o problema de gradientes desvanecentes, mas também permitem que se construam arquiteturas de rede mais profundas e eficazes para

tarefas de visão computacional, alcançando resultados de precisão significativamente melhores em comparação com abordagens anteriores.

Figura 16 – Arquitetura da CNN ResNet



Fonte: He et al. (2016)

2.5 Estimativa Volumétrica de resíduos

A estimativa de volume utilizando visão computacional ainda é um conceito sem um estado da arte completamente definido, conforme levantamento realizado até o presente momento. Esse processo não tem sido um tema de pesquisa amplamente explorado, sendo mais comum em aplicações de estimativa de volume de alimentos (LU et al., 2018; LO et al., 2018; OKINDA et al., 2020; YANG et al., 2021). Não está claro se essa limitação se deve à falta de dados representativos para outras aplicações ou ao uso de sensores como o LiDAR (*Light Detection and Ranging*) (LI et al., 2022), amplamente utilizado na indústria.

Existem diversas técnicas, algoritmos e *frameworks* que podem ser utilizados para esse processo, cada um com suas vantagens e desvantagens. A adoção de métodos baseados em visão computacional pode variar conforme a necessidade e a especificidade do problema, destacando-se a importância de pesquisas futuras para aprimorar essas metodologias e expandir suas aplicações.

2.5.1 Estimativa de profundidade

A estimação de profundidade é considerada uma etapa imprescindível no processo de estimação volumétrica de um objeto através de imagem. Isso se dá pelo fato de que precisamos de informações que não são supridas por apenas uma simples imagem 2D do item a ser analisado.

Com isso em mente, diversos artigos utilizam metodologias e algoritmos variados que melhor se adaptam à realidade necessária para a resolução do problema proposto

(QI et al., 2018; XU et al., 2018; ZHANG et al., 2019). Em geral, as tecnologias mais comumente empregadas no processo de estimação de volume incluem o uso de múltiplas câmeras ou imagens para análise de mudança de posição, técnicas de aprendizado profundo para análise de múltiplos *frames* de um vídeo, ou o uso de câmeras e sensores capazes de obter tais informações (MERTAN; DUFF; UNAL, 2022). Cada abordagem apresenta suas próprias vantagens, desvantagens e desafios específicos.

O processo de estimativa de profundidade é comumente empregado no processo de estimação volumétrica. Onde as abordagens mais comuns utilização um junção de abordagens baseadas em modelos com visão computacional, tais abordagens se fazem necessárias em cenários onde não existem dados representativos do problema a ser solucionado e quando disponíveis, os dados disponíveis não incluem informações de profundidade captadas por lentes e sensores específicos. Isso ocorre porque a aquisição desse tipo de dados é economicamente custosa e um processo complexo, exigindo um grande investimento de recursos. Além disso, dependendo da aplicação, é praticamente impossível encontrar conjuntos de dados prontos para uso (ZHAO et al., 2020; MING et al., 2021).

Dessa forma, alguns autores optam por obter as informações de profundidade a partir de imagens RGB, utilizando um par ou mais de câmeras (LI; HAN, 2018). Apesar de este conjunto de múltiplas imagens conter mais dados e informações 3D, o processo de captura pode ser caro, além de exigir hardware adicional que torna o sistema maior e mais pesado, o que pode levar a problemas como consumo de energia e usabilidade geral.

Autores como Xie, Girshick e Farhadi (2016) e Godard, Aodha e Brostow (2017) utilizaram uma imagem de um par estereoscópico para prever um mapa de disparidade e reconstruir a vista correspondente do objeto analisado, assim obtendo informações de profundidade do objeto.

Com a popularidade e o avanço dos modelos de estimativa de profundidade baseados em aprendizado profundo (MEYERS et al., 2015; LIU; SHEN; LIN, 2015; CHEN et al., 2016), tem sido demonstrado que informações tridimensionais podem ser deduzidas a partir de uma única imagem de vista 3D.

Apesar de esses modelos baseados em aprendizado profundo terem alcançado certo sucesso na melhoria da estimativa de volume usando a estimativa de profundidade, como o im2calories (MEYERS et al., 2015) e o deepvol (LI; HAN, 2018), derivar informações de profundidade a partir de uma única imagem RGB ainda é um problema altamente desafiador. Essa dificuldade se torna evidente ao ser aplicada em ambientes complexos e ao tentar desenvolver um modelo de estimativa mais generalizado. Além disso, esses modelos requerem uma quantidade excessivamente grande de imagens RGB, algumas com informações de profundidade conhecidas, necessárias para treinar a rede, o que coloca desafios práticos na obtenção desses dados.

Dalai, Dalai e Senapati (2023) utilizaram técnicas como SIFT (*Scale Invariant Feature Transform*) e operadores Laplacianos para analisar a forma dos objetos e extrair características da imagem. Essas características foram posteriormente utilizadas pelo *framework* VGG-ResNet para obter dados de profundidade do objeto analisado.

Liao et al. (2021) desenvolveram um método adaptativo de estimativa de profundidade dentro do *framework* Pyramid Multi-View Stereo (MVS). Nos resultados, este *framework*, considerado eficiente em termos de memória, destacou-se na realização de estimação de profundidade de maneira eficiente, oferecendo um ótimo resultado e minimizando o uso de memória do sistema.

De acordo com Xie, Girshick e Farhadi (2016), a estimativa de profundidade monocular tem superado o desafio da limitação de dados de profundidade ao adotar uma abordagem de treinamento auto-supervisionado.

Desenvolvimentos recentes, como observado nos trabalhos de Godard et al. (2019) e Zhou et al. (2017), avançaram na direção de reconstruir objetos através da utilização de imagens passadas e futuras na sequência de vídeo. Essa metodologia utiliza um mapa de profundidade e modelos de identificação de posicionamento da câmera, eliminando assim a necessidade de pré-processamento das imagens na etapa de entrada e permitindo o uso de imagens capturadas com câmeras com sensores comuns, barateando e democratizando o processo de captura e criação do conjunto de dados.

Concluindo, a estimação de profundidade desempenha um papel vital na obtenção de estimativas volumétricas precisas a partir de imagens, superando as limitações inerentes à análise bidimensional. A integração de múltiplas câmeras, a utilização de técnicas avançadas de aprendizado profundo e o desenvolvimento de *frameworks* inovadores, destacam o progresso contínuo e a complexidade desse campo. Embora os desafios, como a necessidade de dados de alta qualidade e o custo elevado de captura, ainda persistam, os avanços metodológicos e tecnológicos estão constantemente expandindo e tornando-a cada vez mais robusta e aplicável em cenários cada vez mais complexos.

2.5.2 Estimativa de volume baseada em imagem

As metodologias, *frameworks* e cálculos responsáveis pela estimação volumétrica dependem diretamente das técnicas, modelos e algoritmos escolhidos para a obtenção e tratamento dos dados que foram explicados anteriormente. Os procedimentos matemáticos e complexidade geral do código vai ser um reflexo dessas escolhas, por isso, é a parte que mais diverge de literatura para literatura.

Podemos categorizar dois principais métodos para estimativa de volume baseada em imagens, são elas a visão computacional e os métodos baseados em modelos, que apesar de parecem diferentes, geralmente são utilizadas em conjunto para um melhor resultado

(LO et al., 2018). Essas estratégias bem estabelecidas têm despertado grande interesse devido à sua ampla gama de aplicações em diversos setores, incluindo educação, saúde e desenvolvimento social em geral.

Cada método possui vantagens e desvantagens, que variam dependendo do modelo proposto e da metodologia utilizada. Em geral, os métodos puramente de visão computacional geralmente requerem múltiplas imagens do objeto-alvo e um objeto de referência de tamanho conhecido para definir a escala de pixel para métrica (JADHAV; SINGH; ABHYANKAR, 2019). Além disso, a implementação da câmera exige que o usuário insira mais informações, o que pode aumentar o risco de erros devido à sua complexidade e comprometer a experiência do usuário (STEINBRENER et al., 2023).

Por outro lado, a abordagem puramente baseada em modelos proporciona uma experiência amigável com mínima intervenção do usuário. No entanto, essa vantagem vem acompanhada da necessidade de dados de alta qualidade, o que pode ser difícil de obter, dependendo da aplicação específica (GRAIKOS et al., 2020; BANDI et al., 2020; DALAI; DALAI; SENAPATI, 2023). Por isso, a utilização das vantagens de cada modelo é a solução mais interessante para esse problema, criando um híbrido que consiga resolver o problema de maneira eficiente, considerando os dados disponíveis.

No grupo de visão computacional, podemos destacar o uso de objetos com dimensões conhecidas, como moedas, para definir um fator de escala de distância pixel-para-métrica, como utilizado por Almaghrabi et al. (2012) e Liang e Li (2017).

O uso de uma placa de calibração xadrez também é muito comum. Xu et al. (2013) utilizam um dicionário de formas de alimentos predefinidas e imagens de alimentos juntamente com uma placa de calibração xadrez para estimar o volume dos alimentos. O sistema então tenta corresponder o objeto alimentar capturado a uma das formas predefinidas no dicionário, cada uma associada a um volume conhecido.

Uma abordagem alternativa dentro deste grupo visa gerar uma representação tridimensional do objeto analisado, utilizando múltiplas imagens. Este modelo 3D serve então como base para a estimativa do volume do objeto. Por exemplo, Puri et al. (2009) propuseram um método que utiliza várias imagens de diferentes pontos de vista para gerar um modelo 3D. De forma semelhante, Hassannejad et al. (2017) empregaram um curto vídeo da refeição do usuário e utilizaram determinados quadros consecutivos para criar uma representação em nuvem de pontos 3D do objeto. Os autores Xu et al. (2013) e Dehais et al. (2017) também utilizaram duas ou mais imagens do objeto analisado para recriar uma perspectiva 3D ou nuvem de pontos.

No entanto, no contexto de todas as metodologias de reconstrução discutidas, a presença de uma placa de calibração xadrez na cena é imperativa. Isso garante a escala e facilita a detecção das transformações de pose da câmera entre as visualizações,

além de destacar a necessidade de múltiplas ações e entradas pelo usuário, o que pode potencialmente levar a impactos negativos nos resultados finais (BANDI et al., 2020).

Para abordar as limitações associadas à interação do usuário e aos requisitos de calibração do grupo de visão computacional, Hassannejad et al. (2017) propuseram uma abordagem baseada em modelos. O método deles utiliza uma rede neural convolucional profunda treinada para prever a profundidade de uma imagem fornecida pelo usuário. Aproveitando os parâmetros intrínsecos da câmera, a rede projeta cada pixel em um ponto 3D correspondente. Essa projeção permite a criação de uma reconstrução 3D aproximada do alimento, similar aos métodos discutidos anteriormente. O volume estimado do alimento é então derivado desse modelo reconstruído.

Desenvolvimentos recentes, como observado nos trabalhos de Godard et al. (2019) e Zhou et al. (2017), têm avançado em direção à reconstrução de imagens passadas e futuras na sequência de vídeo. Este método utiliza um mapa de profundidade previsto e ajuste da posição da câmera, eliminando a necessidade de pré-processamento de imagens na etapa de entrada. Além disso, por meio da perda de treinamento sobre eles e do automonitoramento, essas técnicas permitem o uso de dados capturados por sensores de câmera convencionais.

Em resumo, os métodos de estimativa de volume baseados em modelos frequentemente enfrentam limitações devido ao desafio de obter dados de verdade de terreno de alta qualidade. A estimativa de profundidade da imagem, por exemplo, requer uma captura de imagem de alta fidelidade. Essas limitações impedem os métodos existentes de obter estimativas quantitativas precisas e consistentes (GRAIKOS et al., 2020).

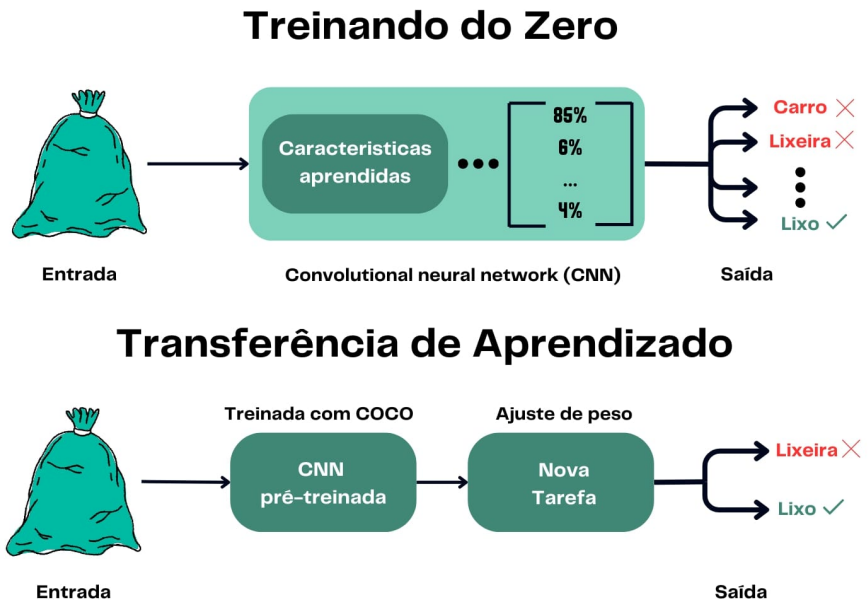
2.6 Transferência de Aprendizado

Embora a tecnologia tradicional de aprendizado de máquina tenha alcançado grande sucesso e sido amplamente aplicada em muitas áreas práticas, ela ainda enfrenta limitações em determinados cenários do mundo real. O cenário ideal para o aprendizado de máquina envolve a disponibilidade de uma grande quantidade de dados de treinamento rotuladas que sejam representativos dos dados de teste e da aplicação real. No entanto, em muitos casos, coletar dados de treinamento suficientes pode ser caro, demorado ou até mesmo impraticável. Além disso, em muitos casos, também é difícil coletar instâncias não rotuladas, o que frequentemente resulta em modelos com resultados insatisfatórios (PADILLA; NETTO; SILVA, 2020).

A transferência de aprendizado, promissora técnica para resolver o problema mencionado anteriormente, pode ser definida como o processo de aprimoramento do desempenho de um modelo utilizando o conhecimento adquirido por meio de uma rede previamente treinada, desde que os domínios sejam relacionados (ZHU et al., 2023). Isso reduz a necessidade de um grande volume de dados para construir modelos eficazes e diminui

o custo computacional. Devido às amplas possibilidades de aplicação, a transferência de aprendizado tornou-se uma área popular e promissora no campo do aprendizado de máquina (PADILLA; NETTO; SILVA, 2020). A Figura 17 mostra a diferença entre o treinamento de um modelo do zero e coma a transferência de aprendizado.

Figura 17 – Diferença de treino de modelo com e sem transferência de aprendizado



Fonte: Elaborado pelo autor

Na figura acima, podemos ver que o modelo que utilizou transferência de aprendizado passou por um ajuste dos seus pesos. Essa etapa é muito importante para que o modelo se adapte melhor e se especialize no cenário específico em que está sendo aplicado. Grandes conjuntos de dados como ImageNet e COCO são essenciais nesse processo de transferência, pois fornecem modelos pré-treinados com milhares de imagens. Esses recursos são amplamente utilizados em diversas aplicações para melhorar e adaptar os modelos, permitindo que reconheçam padrões de forma mais eficaz e se tornem mais especializados (LIN et al., 2014; KRIZHEVSKY; SUTSKEVER; HINTON, 2017).

2.7 Métricas de Avaliação

Para quantificar objetivamente o desempenho dos algoritmos e modelos no sistema, é essencial utilizar métricas de avaliação adequadas. Essas métricas permitem uma análise precisa e comparativa, fornecendo uma base sólida para comparar diferentes abordagens, ajustar hiperparâmetros e validar a eficácia dos modelos no contexto específico do problema. A escolha correta das métricas garante que o modelo atenda aos requisitos e expectativas do projeto, proporcionando resultados confiáveis e relevantes (PADILLA et al., 2021).

A interpretação correta dessas métricas é crucial para uma análise eficaz dos resultados, permitindo uma comunicação clara e fundamentada sobre o desempenho do sistema. A combinação de várias métricas pode proporcionar uma visão mais abrangente e detalhada, assegurando uma avaliação robusta e confiável do desempenho dos modelos (ZHANG; YANG, 2023).

2.7.1 Erro Percentual Relativo (RPE)

O Erro Percentual Relativo, do inglês *Relative Percentage Error*, é uma métrica amplamente utilizada para avaliar o desempenho de modelos preditivos, especialmente em contextos onde a precisão relativa é mais relevante do que a absoluta. No contexto da estimação de volume de resíduos sólidos, o RPE oferece uma medida clara e intuitiva da discrepância entre os valores previstos e os valores reais, expressa como uma porcentagem dos valores reais.

O RPE é definido pela fórmula:

$$\text{RPE} = \left(\frac{|\hat{y} - y|}{|y|} \right) \times 100\% \tag{1}$$

Onde:

- $\hat{y}$ representa o valor previsto pelo modelo.
- y é o valor real observado.

A principal vantagem do RPE é sua capacidade de fornecer uma medida de erro que é dimensionada em relação ao valor real. Isso é particularmente útil em situações onde os valores reais variam amplamente, algo comum no processo de estimação de volume através de imagens. Além disso, como o RPE é expresso em porcentagem, ele facilita a interpretação dos resultados por ser uma métrica intuitiva para a maioria dos usuários.

2.7.2 Erro Percentual Absoluto Médio (MAPE)

O Erro Percentual Absoluto Médio, do inglês *Mean Absolute Percentage Error*), é uma métrica comum para avaliar a precisão de modelos de previsão em termos de porcentagem do erro médio em relação aos valores reais. No contexto da estimação de volume de resíduos sólidos, o MAPE oferece uma medida intuitiva da precisão relativa das previsões e comumente utilizado em literaturas com a problemática de estimação volumétrica, como nos trabalhos de Dehais et al. (2017), Graikos et al. (2020) e Dalai, Dalai e Senapati (2023).

O MAPE é definido pela fórmula:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100\% \tag{2}$$

Onde:

- $\hat{y}_i$ representa o valor previsto para a observação i.
- y_i é o valor real observado para a observação i.
- n é o número total de observações.

O MAPE oferece uma medida média da precisão relativa das previsões, sendo útil para comparar diferentes modelos em termos de sua capacidade de prever com precisão. Ele é especialmente vantajoso por ser intuitivo e escalável em relação aos valores reais, facilitando a interpretação e a comparação de desempenho entre modelos (MYTTENAERE et al., 2016).

Ambas as métricas, RPE e MAPE, são essenciais para avaliar a qualidade das previsões em projetos de modelagem preditiva. Elas permitem uma análise objetiva do desempenho dos modelos, ajudando na escolha do modelo mais adequado e na otimização de hiperparâmetros. A interpretação correta dessas métricas também permite uma comunicação clara sobre a precisão das previsões, crucial para validar a eficácia do modelo.

Ao interpretar o RPE e o MAPE, é importante considerar que valores mais baixos indicam uma menor discrepância entre as previsões e os valores reais, refletindo uma maior precisão do modelo. Essas métricas são sensíveis a *outliers* e variações extremas nos dados, portanto, garantir a consistência e a qualidade dos dados de entrada é fundamental para uma avaliação precisa e confiável do desempenho dos modelos.

2.7.3 Coeficiente de Determinação

Outra técnica comumente utilizada para avaliar modelos de estimação volumétrica é o coeficiente de determinação (R^2, ou R-*squared* em inglês), como demonstrado por Dalai, Dalai e Senapati (2023) em seu trabalho. Introduzido por Wright (1921), o R^2é uma medida estatística que avalia a qualidade do ajuste de um modelo de regressão. Em termos simples, o R^2 indica a proporção da variância da variável dependente que é explicada pelas variáveis independentes do modelo (CHICCO; WARRENS; JURMAN, 2021)

O R^2 é definido pela fórmula:

$$R^2 = 1 - \frac{\sum_{k=1}^{M} |C_k - d_k|^2}{\sum_{k=1}^{M} \left| C_k - \bar{C} \right|^2} \tag{3}$$

Onde:

- R^2 representa a performance do coeficiente de determinação.
- C_k é o valor original, no caso deste trabalho, o volume original.
- d_k é o valor predito pelo modelo.
- $\bar{C}$ é a média dos valores originais.
- M é o número total de observações.

De acordo com Chicco, Warrens e Jurman (2021), o R^2 oferece uma visão sobre o ajuste geral do modelo e sua capacidade de explicar a variância nos dados. Isso é especialmente útil para avaliar a qualidade do modelo de forma geral. Além disso, ao contrário de outras métricas que podem variar de zero ao infinito e que podem não ser intuitivamente fáceis de interpretar, o R-*squared* está normalizado entre 0 e 1 (ou pode ser negativo em alguns casos), o que facilita a interpretação. Valores mais próximos de 1 indicam um modelo melhor, enquanto valores próximos de 0 sugerem um modelo ruim. Além disso, o R^2 tende a ser menos sensível a *outliers*, o que pode proporcionar uma avaliação mais robusta do ajuste do modelo.

2.8 Considerações Finais

Neste capítulo, é apresentada a base teórica que sustenta todos os conceitos e técnicas utilizados neste estudo, com o objetivo de facilitar a compreensão dos capítulos subsequentes. Em seguida, no Capítulo 3, são discutidas as pesquisas relacionadas que tratam de temas similares, destacando como elas contribuíram para a construção deste trabalho e fornecendo uma visão crítica sobre suas abordagens e resultados.

3 Trabalhos Correlatos

3.1 Considerações Iniciais

O presente capítulo tem como objetivo apresentar uma revisão abrangente da literatura e dos trabalhos relacionados aos processos de detecção, segmentação e estimação volumétrica de resíduos sólidos urbanos. A crescente urbanização e o aumento da geração de resíduos sólidos tornaram imperativa a necessidade de soluções eficientes e automatizadas para a gestão de resíduos. Nesse contexto, a visão computacional e as técnicas de *Deep Learning* emergem como ferramentas promissoras.

3.2 Levantamento do Estado da Arte

Com uma metodologia predominantemente estruturalista, conforme discutido por Pereira et al. (2018), e um levantamento da literatura relacionado ao tema que é majoritariamente experimental, conforme definido por Lunetta e Guerra (2023), este estudo fez uso extensivo de diversas fontes. Para o processo de localização das fontes, referências metodológicas e técnicas estudadas e aplicadas nesta pesquisa, foram utilizados diversas bases de dados e bibliotecas virtuais. A Tabela 1 tem como objetivo mostrar as principais bases utilizadas no desenvolvimento desta pesquisa.

Ainda que a estimativa de volume a partir de imagens não seja uma técnica relativamente nova, o campo está em constante evolução e não possui um estado da arte definitivo devido à diversidade de métodos analíticos, à especificidade dos objetos ou cenas observadas e a outras variáveis que introduzem características e requisitos distintos. As abordagens variam desde algoritmos tradicionais de visão computacional até técnicas avançadas de aprendizado profundo, refletem a complexidade da tarefa. Além disso, as diferenças nos dados de entrada, como iluminação e resolução, adicionam desafios adicionais ao processo de estimativa volumétrica. Esse panorama variado promove a inovação contínua e o aprimoramento das técnicas, mas também evidencia a dificuldade em estabelecer uma solução universal que se ajuste a todas as aplicações.

As aplicações de estimação volumétrica são amplas e variadas, frequentemente sendo empregadas em contextos específicos, como a análise de cenas médicas e espaciais, conforme demonstrado por Azarafza, Koçkar e Faramarzi (2021) e Rabbani et al. (2023). No entanto, um dos cenários mais comuns e amplamente aplicado é a estimativa de volume de alimentos. Este domínio se beneficia particularmente da capacidade das técnicas de estimação volumétrica para lidar com a variabilidade na forma e no tamanho dos alimentos,

Tabela 1 – Base dados utilizadas no levantamento da literatura

Base de dados	**Descrição**	**URL**
Scopus	É a maior base de dados de resumos de literatura revisada por pares, com ferramentas bibliométricas para acompanhar, analisar e visualizar a pesquisa. Site da editora Elsevier concentra artigos científicos, revistas e livros. Possui mais de 22000 títulos com mais de 5000 editores em todo mundo, que abrange as áreas de ciências, tecnologia, medicina, ciências sociais, artes e humanidades.	<https://www.elsevier.com/solutions/scopus>
Plataforma Sucupira	É uma ferramenta para coletar informações, realizar análises e ser a base de referência do Sistema Nacional de Pós-Graduação do Brasil, artigos de revistas nacionais e internacionais.	<https://sucupira.capes.gov.br/sucupira/>
Periódicos da Capes	É uma biblioteca da CAPES que contém produções cientifica, tem um acervo de mais 45.000 periódicos completos, 130 bases referenciais, 12 bases dedicadas exclusivamente a patentes, livros, enciclopédias e obras de referência, normas técnicas, estatísticas de conteúdo e audiovisual.	<http://www.periodicos.capes.gov.br/>
Google academic	É uma ferramenta que auxilia na busca de literatura acadêmica como: teses, artigos, livros e outros	<https://scholar.google.com.br/>

Fonte: Adaptado de Sousa, Oliveira e Alves (2021)

permitindo uma avaliação mais precisa e eficiente em diversas aplicações (LO et al., 2020; OKINDA et al., 2020; POPLY; JOTHI, 2021).

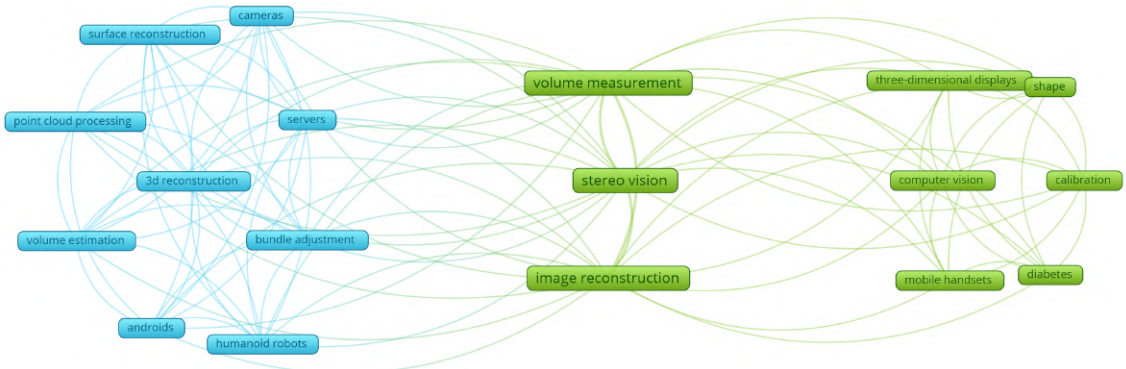
Embora a razão exata para a grande quantidade de trabalhos que analisam alimentos e suas variações ainda não esteja totalmente explicada, pode-se teorizar que a facilidade de obtenção de dados alimentares e a diversidade de análises e aplicações possíveis nesse contexto são fatores significativos. Além disso, a disponibilidade de conjuntos de dados especializados em alimentos, como os desenvolvidos por Kawano e Yanai (2015), Damen et al. (2021), Chen et al. (2017) e Ciocca, Napoletano e Schettini (2016), contribui para o foco elevado nesse tipo de análise. Esses conjuntos de dados facilitam o desenvolvimento e a avaliação de técnicas de estimativa volumétrica, oferecendo uma base rica e variada para experimentação e aprimoramento de métodos, o que pode explicar a prevalência de pesquisas nesse domínio específico.

No levantamento de referências, evitou-se modelos que dependem de sensores adicionais além das câmeras, focando exclusivamente em abordagens baseadas em visão computacional e aprendizado profundo. Embora modelos como os desenvolvidos por Bano, Din e Al-Huqail (2020), Roy et al. (2022) e Gatti, Barbierato e Pozzi (2024) sejam interessantes por tratarem de lixeiras inteligentes capazes de estimar o volume do lixo, eles se limitam a objetos descartados corretamente, o que não reflete a realidade da maioria dos países em desenvolvimento. Além disso, o alto custo desses sistemas restringe sua replicabilidade em países com menos investimento em soluções desse tipo, como o Brasil.

Entre todas as literaturas analisadas, foram filtradas as metodologias que não apresentavam caráter ou possibilidade de serem generalistas. Isso ocorreu porque não foram encontradas tentativas de estimação volumétrica específicas para resíduos sólidos. Portanto, foram utilizadas as abordagens do estado da arte em problemas semelhantes, como a estimação volumétrica de alimentos, para fundamentar as técnicas e modelos a serem empregados no desenvolvimento do sistema proposto. As literaturas filtradas frequentemente se baseavam em conceitos específicos dos objetos analisados, como no caso do estudo de Huynh, TonThat e Dao (2022), que requer que o formato da batata-doce siga um padrão para realizar cortes virtuais para estimar seu volume.

No total, foram analisados 15 artigos entre os anos de 2017 à 2023, sobre estimação volumétrica. Utilizando o *software* VOSviewer, que visa a visualização e análise de dados bibliométricos e é capaz de gerar uma representação gráfica das interligações entre as palavras-chave mais utilizadas nos artigos avaliados (ECK; WALTMAN, 2011), foi gerada a Figura 18, que mostra as interligações mais importantes entre os artigos. Este intervalo de tempo foi escolhido devido à necessidade de compreender a evolução e desenvolvimento do estado da arte na área. A análise revelou que o estado da arte não é definitivo e ainda é escasso, indicando a relevância e a oportunidade de explorar novas abordagens e técnicas na estimação volumétrica de resíduos sólidos.

Figura 18 – Rede bibliométrica de estimação volumétrica



Fonte: Elaborado pelo autor

Com base na Figura 18, é possível constatar que os termos "*volume measurement*", "*stereo vision*" e "*image reconstruction*" são os mais comuns em artigos de estimativa de volume, indicando a diversidade de técnicas aplicáveis a esse processo. A escolha da técnica adequada deve levar em consideração os dados disponíveis, a complexidade e as especificidades de cada problema a ser resolvido. Para um detalhamento mais aprofundado dos trabalhos correlatos, foram selecionados 6 estudos, escolhidos a partir de uma lista inicial de 15, as literaturas datam entre os anos de 2017 à 2023. A seleção foi feita com base na relevância ao tema da pesquisa e na metodologia utilizada.

Em seu artigo, Bandi et al. (2020) desenvolveu um sistema baseado em visão estéreo para estimar o volume de objetos comuns. O sistema proposto utiliza o processo de reconstrução 3D do objeto a partir de múltiplas imagens, onde as posições das câmeras são estimadas com base nos pontos de características encontrados nas imagens. A partir dessas estimativas, é gerada uma nuvem de pontos densa, cujo volume da superfície final é calculado e convertido para escala métrica utilizando um objeto de referência. De acordo com os autores, o sistema alcançou uma acurácia de 90% nos testes, no entanto, é necessário um mínimo de 30 imagens de um objeto para obter esses resultados, o que pode ser considerado uma grande limitação da implementação.

Iniciando os artigos baseados em alimentos e com uma proposta similar à desenvolvida por Bandi et al. (2020), Dehais et al. (2017) desenvolveu um sistema que utiliza dois celulares para capturar múltiplas imagens de um prato de comida para a reconstrução 3D dos alimentos, também necessitando de um objeto de referência. Embora o sistema necessite apenas de duas imagens para a reconstrução do objeto, há a exigência de uma calibração meticulosa dos celulares e a proximidade do objeto analisado. Além disso, a utilização de objetos de referência, apesar de útil em diferentes processos de estimação volumétrica, se torna inviável no contexto de resíduos sólidos, devido à ausência de um padrão nas cenas analisadas. Os objetos utilizados referência geralmente são pequenos tabuleiros de dama capturados juntos a imagem analisada e com o propósito de facilitar a identificação de informações como a posição da câmera e tamanho do objeto analisado (HASSANNEJAD et al., 2017).

Um dos modelos mais singulares propostos para estimar volume foi desenvolvido por Yang et al. (2021). Este modelo é baseado na imitação do processo humano de estimar volumes e utiliza conjuntos de dados de referência de volumes. Após a análise, a imagem do alimento é submetida a uma classificação de probabilidade volumétrica para estimar uma faixa de valores em mililitros para o volume dos alimentos. No entanto, a necessidade de um conjunto de dados representativos e com volumes disponíveis torna essa abordagem complexa e desafiadora para diversos outros modelos de estimação volumétrica.

Os autores Lo et al. (2018) desenvolveram um sistema de estimação volumétrica a partir de apenas uma imagem do objeto analisado. Utilizando a segmentação pelo

Mask R-CNN e um modelo de *Deep Learning* para a criação de outras vistas do objeto, juntamente com um algoritmo de ponto mais próximo iterativo, do inglês *Iterative Closest Point* (ICP), foi possível recriar o objeto em 3D e usar uma nuvem de pontos para calcular o volume. Apesar do excelente resultado de 93% de acurácia, a proposta requer imagens com informações de profundidade, o que encarece e dificulta a obtenção desses dados, especialmente em problemas onde conjuntos de dados específicos não estão disponíveis. Além disso, a necessidade de um conjunto de dados de objetos reais em 3D para a criação das outras vistas dos objetos representa um desafio. Em contextos como a análise de frutas, essa abordagem é plausível, mas deixa de ser eficiente quando não existem formatos padronizados para os itens analisados, como é o caso dos resíduos sólidos.

O trabalho de Dalai, Dalai e Senapati (2023) é um dos mais interessantes no conceito de utilizar apenas modelos de aprendizado profundo para todo o processo de análise, reconstrução 3D e estimação de volume. A metodologia se baseia em identificar o formato do objeto e analisar as suas características, em seguida, é utilizado o *framework* VGG-ResNet para análise de profundidade, onde é então criado a nuvem de pontos e a estimação de volume é feita pela rede neural *Hybrid 3 DU-GNet*. Com uma acurácia de 98.59% e erro percentual absoluto médio de apenas 6.1%, o sistema se provou muito eficiente. O único problema dessa abordagem são os dados, que assim como a proposta de Lo et al. (2018), é necessário dados de imagem com informações de profundidade, no trabalho foi utilizado o NYC Dataset, com mais de 1.449 imagens RGB com informações de profundidade.

Por fim, temos o sistema desenvolvido por Graikos et al. (2020). Os autores criaram um sistema baseado exclusivamente em modelos de aprendizado profundo, sem a necessidade de múltiplas câmeras, muitas ações do usuário ou dados especiais como profundidade ou modelos 3D. O sistema utiliza o Mask R-CNN com o dataset UNIMIB2016 (CIOCCA; NAPOLETANO; SCHETTINI, 2017) para realizar o processo de segmentação dos alimentos, treina o modelo de estimativa de profundidade utilizando o conjunto de dados EPIC-KITCHENS (DAMEN et al., 2021), e usa nuvens de pontos para calcular o volume dos alimentos. Embora exija diferentes conjuntos de dados, estes são comuns e consistem apenas em imagens segmentadas dos objetos e vídeos com movimentos dos objetos analisados. Apesar dos resultados não serem tão impressionantes em comparação com outras metodologias, variando o erro percentual absoluto médio entre 13.13% e 89.09%, o modelo se mostrou um dos mais generalistas de todos os trabalhos analisados.

Os trabalhos descritos anteriormente destacam o estado atual da arte em relação ao processo de estimação volumétrica por meio de imagens, incluindo conceitos, técnicas, modelos e algoritmos que são de extrema importância para a pesquisa realizada neste trabalho. A Tabela 2 apresenta algumas das abordagens de estimação mais comuns, enquanto a Tabela 3 resume as vantagens e desvantagens dos seis artigos descritos anteriormente.

Tabela 2 – Algumas das mais comuns abordagens para estimativa de volume: baseadas em estereoscopia, baseadas em modelos, baseadas em câmera de profundidade e abordagem de aprendizado profundo

Método	**Estereoscopia**	**Modelos**	**Câmera de profundidade**	**Aprendizado profundo**
Preparação	Calibração da câmera.	Construção da biblioteca de modelos 3D.	Calibração da câmera de profundidade.	Treinamento do modelo.
Procedimento	1. Calibração da câmera. 2. Reconstrução do modelo 3D. 3. Geração de malha 3D e determinação da escala. 4. Estimativa de volume.	1. Reconhecimento de objetos. 2. Seleção do modelo na biblioteca . 3. Registro do modelo por rotação e escalonamento. 4. Determinação do volume por modelos pré-construídos.	1. Construção do mapa de profundidade. 2. Reconstrução do modelo 3D baseada em nuvem de pontos. 3. Estimativa de volume.	1. Imagem RGB única capturada. 2. Estimativa do mapa de profundidade. 3. Reconstrução do modelo 3D. 4. Estimativa de volume

Fonte: Adaptado de Lo et al. (2020)

Cada uma das abordagens apresentadas na Tabela 2 possui suas próprias vantagens, desvantagens e requisitos de preparação e procedimento, que se adaptam melhor conforme a problemática a ser resolvida (LO et al., 2020). No entanto, como destacado por Lo et al. (2020) e Graikos et al. (2020), modelos baseados em estereoscopia necessitam de múltiplas câmeras, enquanto câmeras de profundidade exigem hardware específico e frequentemente caro, o que dificulta a democratização da tecnologia. Além disso, modelos que dependem de bibliotecas e conjuntos de dados de objetos 3D enfrentam desafios relacionados ao custo e à complexidade do hardware necessário, bem como à variabilidade dos itens analisados. Diante dessas limitações, os modelos de aprendizado profundo se destacam como uma solução promissora para os problemas mencionados anteriormente.

Tabela 3 – Trabalhos de pesquisa de alto impacto sobre estimativa de volume publicados entre 2017 e 2023

Autores	**Descrição**	**Erro**	**Destaque**	**Desvantagens**
Dehais et al. (2017)	Reconstrução estereoscópica de duas vistas (Estereoscopia)	MAPE variando entre 8.2% e 9.8%; Avaliando em *datasets* de 45 e 14 pratos, respectivamente	Tempo de processamento rápido e bom desempenho da técnica de extração de pose, em inglês *RANdom SAmple Consensus* (RANSAC)	Problemas com múltiplas imagens, e dificuldades quando a textura do alimento não é óbvia

Lo et al. (2018)	Geração de modelos 3D e nuvem de pontos (Modelo e câmera de profundidade)	6,9% para teste no conjunto de dados com 8 objetos alimentares sintéticos (Erro médio de estimativa de volume)	Capacidade de lidar com itens alimentares de formas irregulares e com áreas ocluídas	Dependência de dados que necessitam de equipamentos específicos e limitados a ambientes laboratoriais.
Bandi et al. (2020)	Reconstrução densa estereoscópica de múltiplas vistas (Estereoscopia)	Média de 10% de erro em condições ideais.	Caráter generalista, podendo ser utilizado em diversos tipos de objetos	Necessidade de pelo menos 30 imagens para reconstrução 3D, uso de objeto de referência e baixa precisão em condições não ideais.
Graikos et al. (2020)	Estimativa do mapa de profundidade através de uma arquitetura CNN seguida de cálculo do volume (Aprendizado profundo)	MAPE variando entre 13.73% e 89.04%	Automatização do processo diminui chance de erros, facilidade de uso e caráter generalista	Complexidade computacional e limitação na precisão, mostrando resultados interessantes mas longe do ótimo
Yang et al. (2021)	Geração de modelos 3D e probabilidade volumétrica (Modelo)	Erro Volumétrico Relativo Médio (mRVE) variando entre 11.60% e 20.10%	Consegue estimar o volume mesmo sem informações do diâmetro do prato	Necessidade de *Dataset* com informações de volume e não consegue reconhecer múltiplas instâncias ou tipos de alimentos
Dalai, Dalai e Senapati (2023)	Estimativa do mapa de profundidade através de uma arquitetura CNN seguida de cálculo do volume (Aprendizado profundo) e reconstrução 3D (Modelo)	MAPE de apenas 6.1%	Alta precisão, método híbrido eficaz para reconstrução 3D e eficiência na extração de características	Generalização limitada devido a baixa eficiência em cenários ou com ruídos e necessidade de dados de treinamento de alta qualidade (câmera de profundidade).

Fonte: Elaborado pelo autor

Apesar de termos analisado diversos artigos, alguns foram excluídos dos principais modelos utilizados no desenvolvimento do sistema proposto devido a dificuldades como complexidade excessiva, falta de generalidade, e ausência ou necessidade de dados específicos. No entanto, esses artigos ainda foram úteis, pois forneceram *insights* valiosos sobre abordagens alternativas e desafios comuns na área. Eles ajudaram a identificar limitações e oportunidades para aprimorar nosso próprio modelo, contribuindo para a definição de estratégias mais eficazes e adequadas ao contexto da nossa pesquisa.

Por exemplo o trabalho de Su et al. (2020), que criaram uma representação volumétrica empregando múltiplas camadas de representação. A combinação do volume de custo captura informações em 3D, e a segmentação 2D do objeto foi utilizada para

alcançar uma síntese abrangente de objetos complexos.

Okinda et al. (2020) elaboraram uma metodologia para estimar o volume de alimentos a partir de imagens. Isso envolveu a utilização de análise de curvatura de contorno e uma abordagem de k-mais próximos para o centro de círculo-M para segmentação de objetos nas imagens.

Lu et al. (2018) abordaram a estimativa de volume como um problema de regressão, prevendo volumes a partir de características 3D implícitas derivadas de informações de profundidade transformadas.

Liao et al. (2021) desenvolveram um método adaptativo de estimativa de profundidade dentro do *framework Pyramid Multi-View Stereo*. Nos resultados, este *framework* se destacou ao realizar uma avaliação eficiente de profundidade em alta resolução, enquanto minimizou o uso de memória.

Em seu artigo, Chen et al. (2020) também utilizaram o *framework* MVS. Eles empregaram o *framework* para reconstruir um modelo de nuvem de pontos 3D utilizando múltiplas visualizações de uma única imagem.

Yu et al. (2021) apresentaram um *framework* MVS em forma de pirâmide de custo consciente do conceito para reconstrução 3D, enfatizando um esquema de inferência de profundidade refinada para mapas densos de alta resolução. A abordagem envolve o cálculo do mapa de profundidade em níveis "grosseiros", seguido por *upsampling* iterativo para obter o melhor mapa de profundidade, incluindo a obtenção de auto-atenção em cada camada.

De acordo com Xie, Girshick e Farhadi (2016), a estimativa de profundidade monocular superou o desafio da limitação de dados verdadeiros de profundidade adotando uma abordagem de treinamento auto-supervisionado.

Em seu artigo, He et al. (2021) introduziram uma abordagem inovadora chamada SOSD-net, que integra a segmentação de objetos e o cálculo de profundidade em imagens. Nesse método, a segmentação semântica de objetos e a estimativa de objetos foram realizadas com distintas restrições geométricas.

Os autores em Xie, Girshick e Farhadi (2016) e Godard, Aodha e Brostow (2017) utilizaram uma imagem de um par estereoscópico para prever um mapa de disparidade e reconstruir a vista correspondente do objeto analisado.

3.3 Considerações Finais

Neste capítulo, foram apresentados os trabalhos correlatos que nortearam o desenvolvimento desta pesquisa e forneceram a base para as decisões e o desenvolvimento dos modelos propostos. A revisão dos estudos existentes permitiu identificar as abordagens

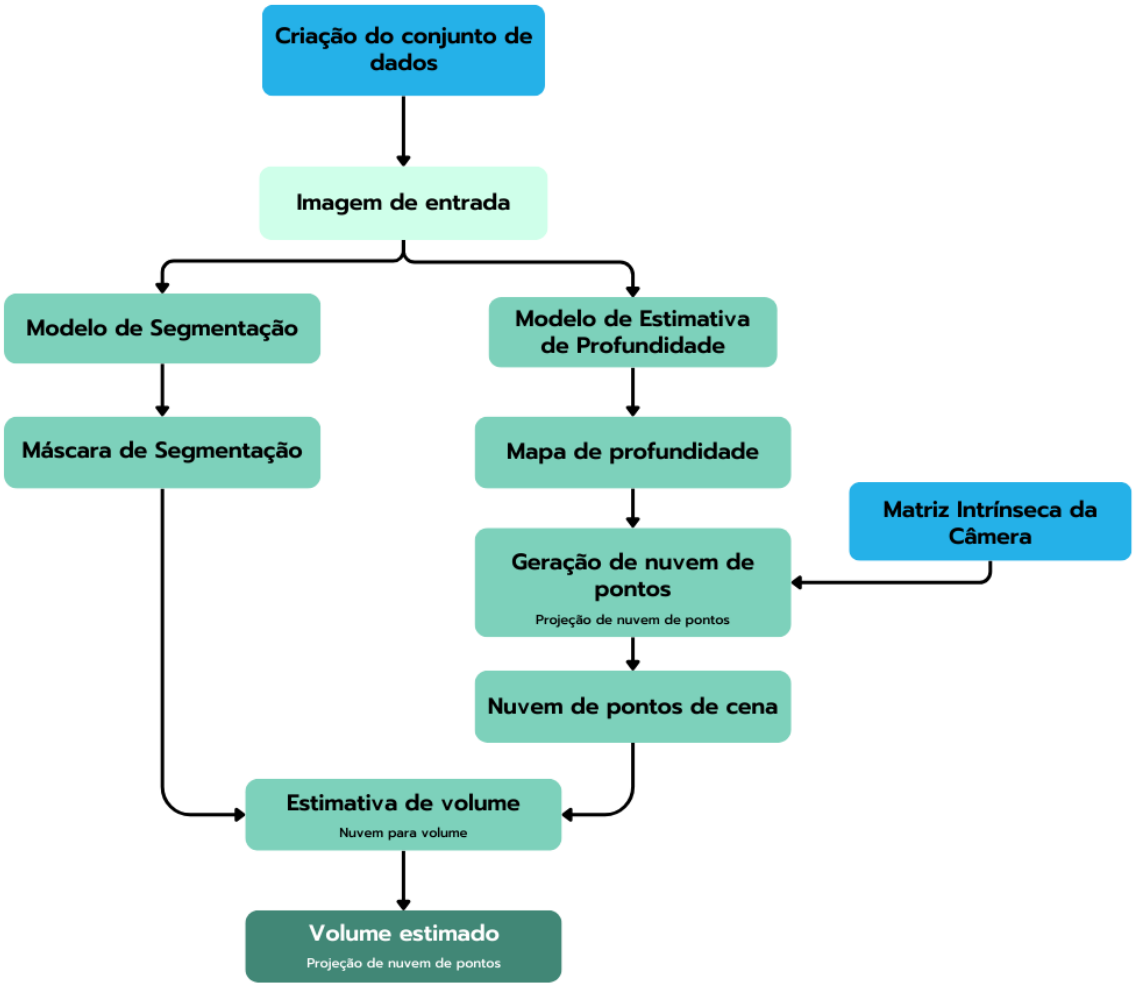
mais eficazes e as lacunas na área de estimação volumétrica por meio de imagens, influenciando a escolha dos algoritmos, *frameworks* e técnicas empregadas. O Capítulo 4 detalha a metodologia desenvolvida, abordando os procedimentos experimentais, a arquitetura dos modelos implementados e a abordagem utilizada na análise dos resultados.

4 Materiais e Métodos

4.1 Considerações Iniciais

Nesta seção, serão detalhadas todas as etapas do desenvolvimento de um sistema para estimação volumétrica de resíduos sólidos a partir de imagens de visão única, utilizando técnicas e modelos de ponta. A *pipeline* do sistema proposto está dividida da seguinte maneira: criação do conjunto de dados (*dataset*), desenvolvimento e treinamento dos modelos de segmentação, estimativa de profundidade e cálculo do volume com base em nuvens de pontos.

Figura 19 – Metodologia proposta para estimativa de volume de resíduos sólidos



Fonte: Adaptado de AZANCORT NETO et al. (2024a)

A Figura 19 ilustra a metodologia final proposta neste trabalho. Após uma extensa

pesquisa em bases de dados públicas de resíduos sólidos urbanos, constatou-se que, devido à especificidade do problema, não foi encontrado nenhum *Dataset* público que fosse útil para a aplicação. Assim, foi criado manualmente um conjunto de dados denominado *WasteInsight* (AZANCORT NETO et al., 2024b).

O processo segue com o processamento da imagem de entrada pelo modelo de segmentação, responsável por criar a máscara de segmentação. Em paralelo, a imagem passa pelo modelo de estimativa de profundidade, que gera o mapa de profundidade. Este, juntamente com os dados da matriz intrínseca da câmera, gera a nuvem de pontos, que é utilizada nos cálculos matemáticos para estimar o volume do resíduo analisado.

Para o treinamento e a avaliação dos modelos de visão computacional utilizados neste estudo, foram empregados um processador Ryzen 5 5600X, uma GPU NVIDIA GeForce RTX 3070 com 8GB de VRAM, 32GB de memória RAM DDR4, e um SSD de 2TB, operando no ambiente Windows 11 com Python 3.6.

No geral, a metodologia proposta neste trabalho tem como objetivo principal democratizar e viabilizar uma solução para a estimativa de volume de resíduos sólidos urbanos sem a necessidade de hardware especializado. Para isso, foi utilizado o modelo de rede neural convolucional Mask R-CNN no processo de segmentação, a arquitetura proposta por Godard et al. (2019) para a estimativa de profundidade e o cálculo de volume proposto por Graikos et al. (2020). Além do conjunto de dados criado especificamente para este trabalho, foi utilizado o conjunto COCO para o treino do modelo de segmentação e o conjunto EPIC-Kitchens para o treino do modelo de estimativa de profundidade.

4.2 Conjuntos de dados

Como comentado anteriormente, não foi encontrado nenhum conjunto de dados público relacionados a RSU, com isso, se fez necessário a criação de um novo *Dataset* representativo para ser utilizado no processo de detecção dos objetos analisados nesta pesquisa. Pensando nisso, foi criado o conjunto público de dados chamado *WasteInsight* que contém 447 imagens divididas em 60% para treinamento, 20% para validação e 20% para testes (AZANCORT NETO et al., 2024b).

As imagens variam em resolução, orientação e condições de iluminação, capturadas em diversos ambientes residenciais, como cozinhas e áreas de serviço. As sacolas de lixo presentes nas imagens apresentam uma variedade de cores, texturas e tamanhos, refletindo a diversidade encontrada em ambientes domésticos reais. Todas as imagens foram capturadas utilizando as câmeras de dispositivos móveis.

Figura 20 – Imagens presentes no *Dataset*

Fonte: Acervo do autor

O conjunto de dados *WasteInsight* é totalmente gratuito e de código aberto e pode ser acessado por meio do Mendeley Data. É esperado que a disponibilidade desses dados ajude a promover a transparência, replicabilidade e colaboração na comunidade científica, facilitando o progresso do conhecimento e a criação de soluções inovadoras na gestão de resíduos sólidos urbanos (AZANCORT NETO et al., 2024b).

O *Dataset* COCO (LIN et al., 2014) também foi utilizado para o treino do processo de segmentação. O modelo treinado utilizou os pesos pré-treinados do conjunto de dados para iniciar o treinamento, ou seja, utilizando a técnica de transferência de aprendizado e depois foi feito o refinamento do modelo com o já comentado *WasteInsight*.

Foi utilizado também o *Dataset* EPIC-KITCHENS (DAMEN et al., 2021) no processo de treinamento do modelo de estimativa de profundidade. Apesar de originalmente ser um conjunto de dados voltado para comida e utilizado em artigos como do Graikos et al. (2020), o treino do modelo mostrou-se muito promissor e com ótimos resultados mesmo em aplicações totalmente diferentes, como a da proposta desse trabalho. O conjunto de dados é composto por mais de cinquenta horas de vídeos egocêntricos que capturam atividades de manipulação de alimentos, resultando em um total de 42,066 *frames*.

4.3 Anotação e Pré-processamento

O processo de anotação para dos dados do treino da segmentação foi todo feito de maneira manual para todas as imagens do conjunto de dados *WasteInsight*. Para esse processo, foi utilizado o *software* VGG Image Annotator (VIA) (DUTTA; GUPTA; ZISSERMANN, 2016). O VIA é um projeto de código aberto baseado exclusivamente em HTML, Javascript e CSS (sem dependência de bibliotecas externas). Desenvolvido pelo Visual Geometry Group e lançado sob a licença BSD-2-Clause, o VIA é útil tanto para projetos acadêmicos quanto para aplicações comerciais.

Foi utilizada a opção de anotação com polilinha, onde foram desenhados contornos lineares ao redor dos resíduos presentes no conjunto de dados. A partir dessas anotações, o programa gera informações sobre o arquivo, incluindo o tipo de objeto analisado, que, para este problema, foi classificado como "saco plástico". Além disso, todos os pontos marcados nos eixos X e Y representam as bordas dos objetos analisados. A figura Figura 26 mostra exemplos dos resultados da segmentação gerada pelo programa.

Figura 21 – Imagens segmentadas presentes no *Dataset*

Fonte: Acervo do autor

4.4 Desenvolvimento dos modelos

4.4.1 Segmentação

O modelo escolhido para a rede de segmentação de instâncias foi o Mask R-CNN (HE et al., 2017), uma extensão da Fast Region-based CNN (GIRSHICK, 2015), utilizada para detecção de objetos. Esse *framework* é capaz de localizar múltiplos objetos de resíduos sólidos presentes em uma imagem e prever uma máscara de segmentação individual para cada instância, permitindo estimar o volume de cada instância separadamente. Embora não tenha sido utilizado neste trabalho, o modelo é capaz de discernir entre os diferentes classes e tipos de resíduos sólidos presentes em uma imagem.

Figura 22 – Exemplos únicos (a) e múltiplos (b) de resíduos sólidos contidos no conjunto de dados coletados

(a) Único Resíduo

(b) Múltiplos Resíduos

Fonte: Acervo do autor

Assim como feito por Graikos et al. (2020), que utilizaram os pesos do conjunto de dados COCO para sua rede de segmentação, neste trabalho adotou-se uma abordagem semelhante, inicializando nosso modelo com esses pesos. No entanto, nosso processo de refinamento envolveu o ajuste fino da rede utilizando um conjunto de dados distinto (*WasteInsight*), composto por objetos de resíduos sólidos urbanos, uma vez que não conseguimos encontrar um conjunto de dados utilizável e pronto para nossa aplicação (AZANCORT NETO et al., 2024b).

O conjunto de dados é composto por imagens de resíduos sólidos em sacolas plásticas, uma vez que esta é a forma mais comum de descarte de lixo no Brasil. O conjunto de dados é bastante simples, contendo apenas as anotações dos objetos para treinamento do modelo. Como foi utilizado apenas um tipo de resíduo sólido, não foram utilizados rótulos ou tipos de objetos, exceto um nome genérico aplicado a todos, denominado "plástico", já que sacolas plásticas foi o único objeto analisado (AZANCORT NETO et al., 2024b).

Para o treinamento da rede de segmentação usando o algoritmo Mask R-CNN, foi utilizado o mesmo *batch size*, taxa de aprendizado (*learning rate* e as mesmas augmentações de dados aplicadas na rede de estimativa de profundidade. Os valores desses e de outros parâmetros utilizados podem ser encontrados na Tabela 4. O modelo foi então treinado com os pesos pré-treinados do conjunto de dados COCO, e ajustado utilizando uma divisão já comentada anteriormente de 60% imagens para treinamento, 20% para validação e 20% para teste da estimativa de volume.

Tabela 4 – Parâmetros de treinamento do modelo de segmentação

Parâmetro	**Valor**
Batch Size	2
Detection Min Confidence	0.7
Detection NMS Threshold	0.3
Learning Rate	1e-3
Validation steps	51
Backbone	resnet101
Training epochs	10

Fonte: Elaborado pelo autor

4.4.2 Estimativa de profundidade

Este trabalho adota a arquitetura de rede introduzida por Godard et al. (2019) para a tarefa de estimativa de profundidade. Em seu trabalho, a rede de estimativa de profundidade do autor foi treinada exclusivamente com sequências de vídeo monoculares, onde cada etapa envolve a utilização de três quadros consecutivos do vídeo (I_{T-1}, I_T, I_{T+1}) no processo de treinamento. A rede de estimativa de profundidade gera um mapa de profundidade, ou *Depth Map* (D_T), para o quadro de entrada (I_T). Simultaneamente, uma rede de estimativa de pose produz as transformações de pose da câmera ($T_{t\to t-1}, T_{t\to t+1}$), representando as relações entre o quadro atual e seus quadros adjacentes. Utilizando o mapa de profundidade, as transformações de pose e a matriz intrínseca conhecida da câmera (K), quadro central é reconstruído a partir de amostras dos quadros anterior e posterior.

$$\begin{aligned} I_{t-1\to t} &= I_{t-1} \left\langle proj(D_t, T_{t\to t-1}, K) \right\rangle \\ I_{t+1\to t} &= I_{t+1} \left\langle proj(D_t, T_{t\to t+1}, K) \right\rangle \end{aligned} \tag{1}$$

Aqui, "proj"refere-se ao método de projeção de coordenadas detalhado por Zhou et al. (2017), e "$\langle\ \rangle$"denota o operador de amostragem. O *Training Loss* compreende a soma de uma perda fotométrica, ou *Photometric Loss* (L_p), que mede a disparidade entre as imagens sintetizadas e originais, juntamente com um erro de suavidade de profundidade, ou *Depth Smoothness Error* (L_s). Este erro de suavidade de profundidade é expresso como uma função que avalia a suavidade do mapa de profundidade previsto (*Predicted Depth Map*).

$$L = L_p + \lambda L_s \tag{2}$$

Essa abordagem, apesar de eficiente no ambiente ideal, encontra dificuldades em produzir um sinal de treinamento significativo quando as transformações de pose são nulas. Nesses casos, os valores de profundidade previstos não têm impacto no processo de síntese de imagem. Essa restrição reduz a seleção de vídeos adequados para treinar a rede àqueles

que apresentam movimento substancial entre os quadros. Essa limitação pode ser difícil de superar, uma vez que vídeos de resíduos sólidos podem apresentar movimento inerente limitado, dependendo de como a captura está sendo feita, tornando desafiador garantir um sinal de treinamento robusto para a rede, já que a ausência de movimento significativo entre os quadros prejudica a eficácia dos mecanismos de treinamento.

Essa limitação nos impediu de treinar nosso próprio modelo. Como mencionado anteriormente, conjuntos de dados adequados para essa tarefa são escassos. Coletar esses dados não é apenas difícil e demorado, mas também pode representar riscos à saúde se manuseados por pessoal não qualificado.

Foi utilizado o modelo e os pesos desenvolvidos por Godard et al. (2019) para a estimativa de profundidade. Embora o artigo original tenha aplicado o modelo especificamente para a estimativa de volume de alimentos, os testes realizados demonstraram que o modelo treinado com o conjunto de dados EPIC-KITCHENS (DAMEN et al., 2021), que inclui mais de cinquenta horas de vídeos egocêntricos de atividades de manipulação de alimentos e posteriormente ajustado pelos autores com 38 vídeos capturados por câmeras de *smartphones* comerciais, apresentou resultados promissores ao comparar a estimativa de volume com os volumes conhecidos de objetos de resíduos sólidos urbanos.

A Tabela 5 demonstra todos os parâmetros utilizados no processo de treinamento do modelo de estimativa de profundidade.

Tabela 5 – Parâmetros de treinamento do modelo de estimativa de profundidade

Parâmetro	**Valor**
Input Resolution	128x224
Depth Outputs Range	0.01 to 10
Smoothness Term	10e-2
Training Epochs	20
Learning Rate	10e-4
Ground Truth Expected Median Depth	0.50

Fonte: Elaborado pelo autor

4.4.3 Estimativa de volume

Após a exploração de diversos modelos, algoritmos e técnicas de estimativa de volume em trabalhos anteriores, este artigo adota o método apresentado por Graikos et al. (2020). Conforme proposto pelos autores, utilizamos o mapa de profundidade (D) extraído da imagem de entrada e a matriz intrínseca da câmera (K) para projetar cada pixel (x, y) em seu ponto correspondente no espaço 3D. Essa projeção é realizada utilizando coordenadas homogêneas e o modelo de projeção inversa, resultando em uma representação de nuvem de pontos (P).

$$P_{xy} = K^{-1} \left[\, x \; y \; 1 \, \right]^T D_{xy} \tag{3}$$

Para aprimorar a diferenciação entre vários objetos de resíduos sólidos e particionar o conjunto (P) em subconjuntos distintos de pontos de resíduos sólidos, foi utilizada a máscara de segmentação gerada pela rede de segmentação de instâncias. O pré-processamento de cada conjunto de pontos inicia-se com a remoção de valores atípicos por meio de um filtro de remoção de valores atípicos estatísticos, conhecido como *Statistical Outlier Removal Filter* (SOR). Em seguida, emprega-se a análise de componentes principais, ou *Principal Component Analysis* (PCA), para identificar o plano primário em que o objeto analisado está localizado. Após a aplicação do PCA, o autovetor correspondente ao menor autovalor é selecionado para representar o vetor normal do plano de base no qual o objeto se encontra. Para garantir consistência na orientação do objeto, um passo adicional é implementado para assegurar que o plano esteja posicionado na parte inferior do objeto.

Embora Graikos et al. (2020), que propuseram essa metodologia, tenham utilizado um prato plano para a estimativa de volume, foi adotado o solo como base plana nesta abordagem, desconsiderando a sugestão original. Embora essa escolha possa ser aprimorada, a implementação de uma base plana não é trivial, devido à variabilidade no descarte inadequado de resíduos. Posteriormente, a nuvem de pontos projetados é utilizada para construir um α-complexo derivado da triangulação de Delaunay (EDELSBRUNNER; HARER, 2022). O volume estimado é então definido pela média de cada vértice dos triângulos que compõem o objeto de resíduos sólidos analisado.

Dado que os vídeos utilizados para o treinamento da rede de profundidade não possuem informações verdadeiras de profundidade, as estimativas de profundidade não estão em uma escala métrica. Para lidar com a ausência de informação no processo de treinamento, adotamos a técnica de reescalonamento da mediana da verdadeira escala, também chamada de *Median Ground Truth Rescaling*, proposta por Zhou et al. (2017). Este método dimensiona o mapa de profundidade previsto (D) por um fator constante.

$$s = \frac{mediana(D^{gt})}{mediana(D)} \tag{4}$$

Nos experimentos, o fator de escala é determinado pela mediana da profundidade verdadeira, ou *Median Ground Truth Depth* ($mediana(D^{gt})$), que aproxima a distância entre o sensor da câmera e o objeto analisado. Para os testes, foi assumido um valor de 0,5 metros para todos os casos. Embora esse valor possa ser ajustado manualmente conforme o cenário e a distância do objeto, ele foi fixado para simplificação. Na inferência de volume, o valor do ângulo de campo de visão do sensor da câmera foi ajustado de 70° para 79,5°, o que afeta os dados intrínsecos da câmera. Os valores do Z-Score para o filtro SOR e do α-complexo não foram alterados.

4.5 Avaliação de Performance

A avaliação de desempenho dos modelos foi realizada utilizando as métricas descritas na seção 2.7. Para mensurar a acurácia do sistema no processo de estimação de volume, foi empregado o MAPE e o RPE. Essas métricas são amplamente utilizadas em problemas de estimação de volume, conforme evidenciado na literatura (DEHAIS et al., 2017; STEINBRENER et al., 2023; DALAI; DALAI; SENAPATI, 2023).

Durante a avaliação do modelo treinado para segmentação, foram considerados diferentes valores de perda para analisar a performance. Essa abordagem permitiu observar como as variações na função de perda influenciam a qualidade da segmentação obtida. Os resultados detalhados dessas análises, incluindo a comparação entre os diferentes valores de perda, estão apresentados na Figura 23.

Através dessas métricas e análises, buscamos garantir que o modelo não apenas segmente corretamente os resíduos sólidos, mas também estime com precisão seus volumes.

4.6 Considerações Finais

Neste capítulo, foi realizada a análise e descrição dos elementos, arquitetura, técnicas, algoritmos, dados e fórmulas matemáticas necessárias para estimar o volume de resíduos sólidos urbanos. Com a descrição detalhada do processo de fluxo de desenvolvimento, treino e funcionamento de cada etapa, espera-se que a solução proposta para o problema apresentado neste trabalho seja transparente e facilite a replicabilidade para a criação de soluções inovadoras, não apenas na gestão de resíduos, mas também em diversas áreas que possam se beneficiar da estimativa volumétrica através de imagens. O Capítulo 5 apresenta os resultados obtidos e a análise de desempenho dos processos descritos neste capítulo.

5 Resultados e Discussão

5.1 Considerações Iniciais

Este capítulo apresenta os resultados obtidos através da metodologia proposta e desenvolvida neste trabalho para a estimativa de volume de resíduos sólidos urbanos. Os resultados são fruto do treinamento, testes e validações dos modelos desenvolvidos, e com o conjunto de dados elaborado. As métricas utilizadas para avaliar o desempenho dos modelos foram aquelas detalhadas na seção 2.7.

5.2 Segmentação

Conforme mencionado anteriormente, a segmentação das imagens foi realizada usando o modelo Mask R-CNN, treinado com o conjunto de dados *WasteInsight* dividido em 60% para treinamento, 20% para teste e 20% para validação (AZANCORT NETO et al., 2024b). O modelo foi avaliado com base nos valores das funções de perda durante o processo de treinamento e validação. Além da perda total do modelo, foram analisadas as métricas de perda da função de classificação da Rede de Proposição de Regiões e da função de classificação do Módulo de Regressão e Classificação, *Mask Regional Convolutional Neural Network* (MRCNN).

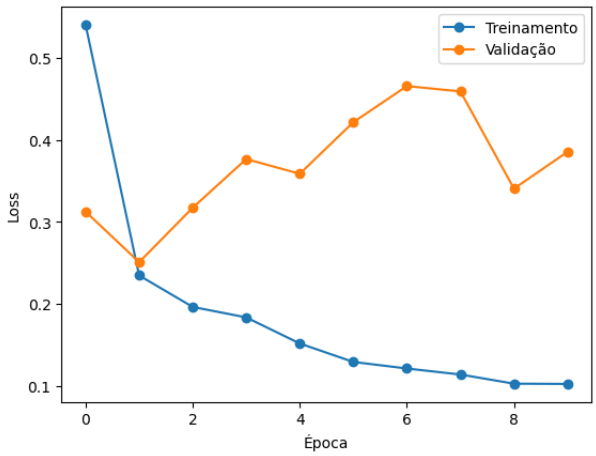
Com base nos resultados das métricas de perda, podemos inferir que o processo de transferência de aprendizado do conjunto de dados COCO foi eficaz para proporcionar a generalização necessária.

Mesmo com um conjunto limitado de dados, é possível observar na Figura 23 que os resultados das funções de perda no processo de treino e validação mostraram um bom resultado. O Modelo foi treinado com 10 épocas, onde cada uma levou um tempo estimado de 4 horas de execução. Essa demora pode ser justificada pelo poder computacional necessário para rodar o modelo com um *batch size* maior que o utilizado.

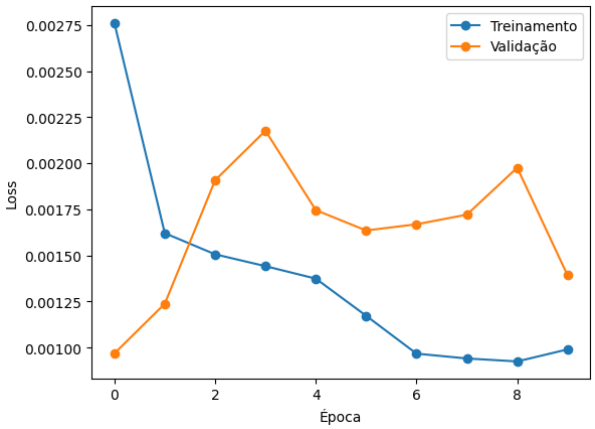
O aumento de dados com transformações no brilho, contraste, saturação, matriz e *flips* horizontais e verticais nas imagens também adicionam tempo de processamento. Por fim, a falta de utilização de GPU foi um fator limitante, já que o treinamento dependeu exclusivamente do processamento da CPU, resultando em tempos de treinamento consideravelmente mais longos do que o esperado.

Figura 23 – Resultados de Loss obtidos no modelo de segmentação

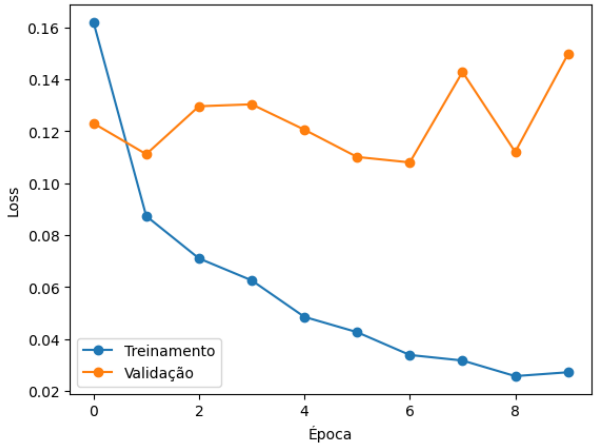
(a) Loss - Perda Total do Modelo



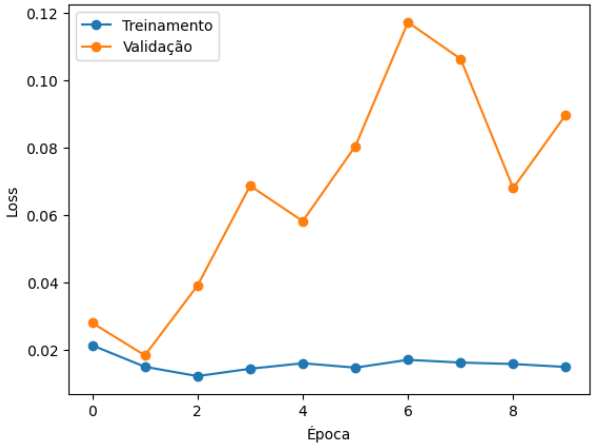
(b) Loss - Classificação da RPN



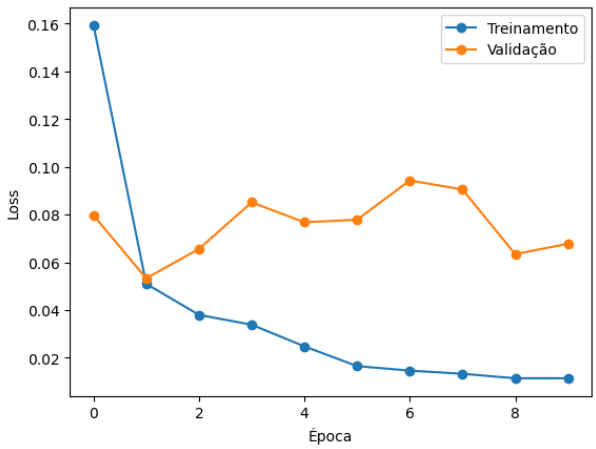
(c) Loss - Regressão de Caixa Delimitadora da RPN



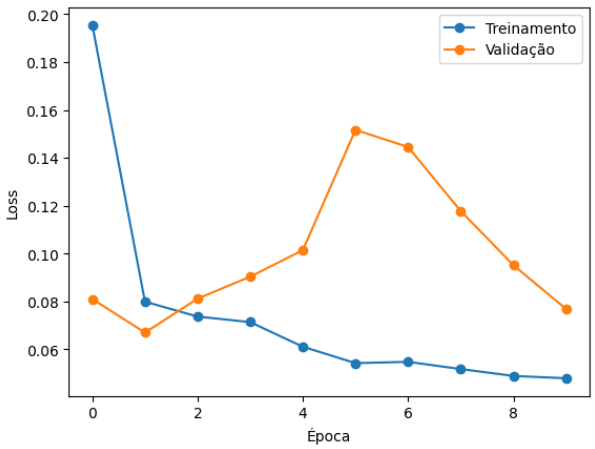
(d) Loss - Classificação do MRCNN



(e) Loss - Regressão de Caixa Delimitadora do MRCNN



(f) Loss - Segmentação de Máscara do MRCNN



Fonte: Acervo do autor

A perda total do modelo, mostrada na Figura 23 (a), é uma métrica composta que soma todas as perdas específicas do modelo Mask R-CNN. Ela representa uma visão

geral da performance do modelo, englobando todos os aspectos do aprendizado. Durante o treinamento, a perda total do modelo apresentou uma tendência de decrescimento consistente, iniciando em 0.5405 e reduzindo-se para 0.1024 na última época, indicando uma melhora significativa na capacidade do modelo de aprender os padrões dos dados de treinamento. A perda total de validação variou ao longo das épocas, começando em 0.3125 e alcançando 0.3854 na última época, o que sugere uma boa capacidade de generalização, apesar de algumas flutuações.

A Rede de Proposição de Regiões é responsável por gerar propostas de regiões que potencialmente contêm objetos. A perda de classificação da RPN mede a precisão da rede ao distinguir entre regiões que contêm objetos e aquelas que não contêm. Durante o treinamento, a perda de classificação da RPN foi consistentemente baixa, diminuindo de 0.0028 para 0.0009, como indicado na Figura 23 (b), indicando que o modelo está melhorando continuamente na distinção de regiões relevantes. A perda de validação da RPN, mantendo-se em torno de 0.001, com uma leve variação, sugere que a capacidade de classificação da RPN é estável e eficaz em dados não vistos.

A perda de regressão de caixa delimitadora da RPN avalia a precisão das coordenadas das caixas delimitadoras geradas para os objetos nas propostas de regiões. Uma perda menor indica que o modelo está predizendo com mais precisão as posições e tamanhos dos objetos. Durante o treinamento, essa perda diminuiu significativamente, como demonstrada na Figura 23 (c), iniciando em 0.1619 e finalizando em 0.0271, indicando uma melhora substancial na precisão das caixas delimitadoras. A perda de validação apresentou variações, com um pico em 0.1496 na nona época, mas geralmente manteve uma tendência de estabilização, sugerindo uma boa capacidade de generalização.

O módulo MRCNN, demonstrado na Figura 23 (d), é responsável pela classificação das regiões propostas pela RPN nas categorias específicas de objetos. A perda de classificação do MRCNN mede a precisão dessa classificação. Durante o treinamento, essa perda diminuiu de 0.0212 para 0.0148, indicando que o modelo está melhorando na classificação correta dos objetos. A perda de validação apresentou algumas flutuações, especialmente nas últimas épocas, indo de 0.0279 para 0.0897, mas ainda assim, esses valores estão dentro de uma faixa aceitável e indicam que o modelo é capaz de aprender de forma eficaz.

A perda de regressão de caixa delimitadora do MRCNN, Figura 23 (e), avalia a precisão das caixas delimitadoras refinadas para os objetos detectados. Durante o treinamento, essa perda diminuiu de 0.1592 para 0.0114, sugerindo que o modelo está se tornando muito mais preciso na predição das caixas delimitadoras. A perda de validação mostrou flutuações, mas uma tendência geral de estabilização em torno de 0.0677 na última época, o que indica uma capacidade aceitável de generalização, embora com alguma variação que pode ser ajustada com mais treinamento ou técnicas de regularização.

A perda de segmentação de máscara do MRCNN, demostrado na Figura 23 (f),

mede a precisão com que o modelo segmenta os objetos dentro das caixas delimitadoras detectadas. Essa métrica é crucial para tarefas de segmentação semântica, onde a precisão de pixel é importante. Durante o treinamento, a perda diminuiu de 0.1954 para 0.0479, indicando uma melhoria significativa na capacidade de segmentação do modelo. A perda de validação apresentou variação ao longo das épocas, mas terminou em um valor similar ao inicial, em 0.0769, sugerindo que o modelo ainda pode estar enfrentando desafios de generalização na tarefa de segmentação, mas com um desempenho geral positivo.

No geral, os gráficos das funções de perda mostrados na Figura 23, indicam que o modelo está aprendendo de maneira consistente durante o treinamento, com todas as métricas de perda apresentando uma tendência decrescente. As métricas de validação, embora mostrem algumas variações, estão em uma faixa aceitável e indicam que o modelo possui uma boa capacidade de generalização. É importante destacar que a quantidade de dados de validação pode ser um dos fatores que influenciam essas flutuações. Com mais dados de validação, o modelo pode ter uma base mais robusta para demonstrar seu desempenho de generalização. A otimização do código para uso de GPU é uma próxima etapa crucial que poderá acelerar significativamente o processo de treinamento, permitindo treinos mais rápidos e a realização de mais testes com diferentes divisões do conjunto de dados e diferentes parâmetros, o que potencialmente melhorará ainda mais a performance do modelo.

5.3 Estimativa de Volume

O processo de estimativa de volume é uma junção de todos os modelos, técnicas e formulações matemáticas mencionadas anteriormente no Capítulo 4. Para a avaliação do sistema proposto, foi realizada a medição de 20 sacos plásticos, que serão denominados "Resíduo Sólido". Cada resíduo utilizado no processo de teste teve múltiplas imagens capturadas de diferentes ângulos e distâncias em relação ao objeto analisado, embora tenha sido tentado manter um padrão na distância da câmera para o objeto.

Para a medição real dos objetos, todos os sacos plásticos foram calculados, seja como instâncias únicas ou múltiplas, utilizando uma representação de polígono irregular. Embora essa abordagem não ofereça uma precisão exata dos valores dos objetos analisados, isso é reconhecido como um dos aspectos a serem aprimorados futuramente. No entanto, é uma estimativa próxima do valor real.

Considerando os variados formatos irregulares, não é possível definir facilmente uma forma de cálculo preciso do volume. A dificuldade reside na complexidade e na variabilidade das formas dos sacos plásticos, o que torna desafiador aplicar um modelo matemático único e preciso para todas as situações. Alguns trabalhos utilizam métodos como o deslocamento de água; contudo, essa abordagem não só não representa com precisão o valor real do

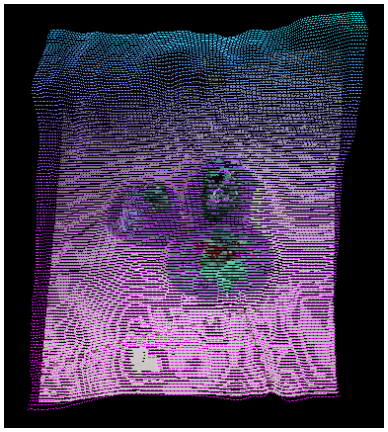
objeto em análise, mas também não é adequada para o tipo de objetos analisados.

Para tornar mais evidente o item analisado no processo final de estimação volumétrica, o sistema não apenas retorna o resultado em litros do volume estimado, mas também gera a visualização da imagem com as nuvens de pontos. Essa visualização permite modificar o tamanho dos pontos para destacar o item analisado, além dos pontos que diferenciam o item do fundo, como exemplificado na Figura 24.

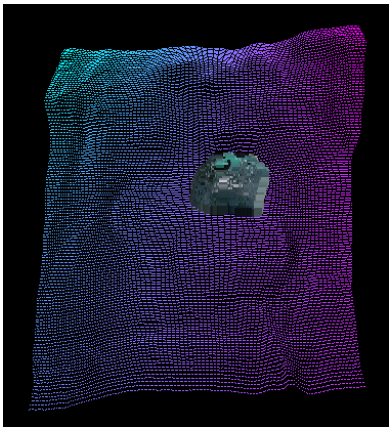
Figura 24 – Representação de nuvem de pontos

(a) Imagem original

(b) Nuvem de pontos de todos os itens



(c) Nuvem de pontos de um dos itens estimados



Fonte: Acervo do autor

Utilizando as métricas RPE e MAPE descritas na seção 2.7, foi possível analisar os resultados do modelo proposto no processo de estimativa de volume de instâncias únicas e múltiplas de resíduos sólidos.

5.3.1 Objetos Individuais

Os objetos individuais são definidos como imagens que mostram apenas uma única instância do resíduo sólido a ser analisado, como mostrado na Figura 25. Conforme discutido anteriormente, variando a angulação e a distância da câmera para cada imagem capturada do objeto analisado. Em nossos testes mais comuns, foram capturadas duas imagens de cada objeto com essas variações.

Figura 25 – Exemplos de Resíduos Sólidos Únicos

Fonte: Acervo do autor

Os resultados de todos os testes são expressos em litros, com um Campo de Visão (FOV) de 79.5° e um estimador de disância com uma variação de 0.01 a 10 metros. Esses valores representam, respectivamente, a distância mínima e máxima do objeto a ser analisado.

Tabela 6 – Volume real medido, o RPE de cada variação e o MAPE estimado a partir de instâncias únicas de resíduos sólidos

Tipo	**Volume Real (L)**	**RPE 1**	**RPE 2**	**MAPE**
Resíduo Sólido Único 1	0.9240	1.11%	0.74%	0.92%
Resíduo Sólido Único 2	0.6555	0.59%	7.08%	3.84%
Resíduo Sólido Único 3	0.7744	3.35%	9.85%	6.60%
Resíduo Sólido Único 4	0.9936	10.39%	2.83%	6.61%
Resíduo Sólido Único 5	0.8008	0.27%	14.73%	7.50%
Resíduo Sólido Único 6	1.6128	10.91%	7.87%	9.39%
Resíduo Sólido Único 7	0.5440	11.29%	9.52%	10.40%
Resíduo Sólido Único 8	1.3000	15.25%	9.32%	12.29%
Resíduo Sólido Único 9	1.0192	16.39%	9.59%	12.99%
Resíduo Sólido Único 10	0.7200	3.33%	27.68%	15.50%

Fonte: Elaborado pelo autor

A partir da análise dos resultados apresentados na Tabela 6, podemos avaliar o desempenho da estrutura proposta em casos onde apenas o volume de um único item está estimado. Como mencionado no Capítulo 3, não foi encontrado outros estudos que utilizem visão computacional para estimar volumes de resíduos sólidos através de imagens. Dessa forma, foi realizado uma comparação direta com os artigos de Graikos et al. (2020), Bandi et al. (2020) e Dalai, Dalai e Senapati (2023) e os respectivos resultados, onde apesar da diferença entre os cenários de aplicação, apresentam semelhanças na sua metodologia.

Em comparação direta utilizando as métricas de desempenho, foi observado que, apesar do conjunto de dados pequeno e limitado, o MAPE de instâncias únicas apresentou uma média de 8.60%, valor significativamente inferior aos 28.40% dos 10 melhores resultados do trabalho de Graikos et al. (2020), ficou equiparado aos 8.2% do trabalho de Bandi et al. (2020) e um pouco maior em comparação ao trabalho de Dalai, Dalai e Senapati (2023) com 6.1%.

Os resultados mostraram valores máximos de 15.50% e mínimos de 0.92%, enquanto o estudo de Graikos et al. (2020) apresentou um valor máximo de 108.3% e mínimo de 13.73% e os outros trabalhos não apresentaram tais informações. Isso demonstra que o modelo proposto, apesar das similaridades com os dos autores mencionados, mostrou-se mais consistente, em alguns casos, com valores de erro consideravelmente menores.

Apesar da eficiência e das vantagens em relação a outras métricas de comparação, o coeficiente de determinação ainda é pouco utilizado em literaturas sobre estimação de volume. Dos artigos analisados, apenas os autores Dalai, Dalai e Senapati (2023) utilizaram essa métrica em seu trabalho. Utilizando a Equação 3 com os dados da Tabela 6, foi obtido um R^2 de 95.11% para resíduos únicos, um resultado bastante satisfatório e próximo do estado da arte. Em comparação, Dalai, Dalai e Senapati (2023) alcançaram um excelente coeficiente de determinação de 98.2

5.3.2 Múltiplos Objetos

Podemos definir múltiplos objetos como cenários onde a imagem apresenta diversas instâncias do objeto a ser analisado. Casos de múltiplas instâncias são mais suscetíveis a erros de estimativa, pois a sobreposição de resíduos aumenta a possibilidade de falhas no processo de segmentação, na estimativa de profundidade e, consequentemente, no cálculo final do volume.

Figura 26 – Exemplos de Múltiplos Resíduos Sólidos

Fonte: Acervo do autor

Para mitigar esses erros, durante o processo de segmentação do conjunto de dados, adotou-se a abordagem de segmentar os resíduos como um único item nos casos em que não era possível distinguir claramente as bordas de um resíduo em relação ao outro. Embora essa abordagem resulte na perda da vantagem do modelo desenvolvido de estimar cada item de forma individual, ela proporcionou resultados promissores. Essa estratégia mostrou-se eficaz em minimizar os erros de sobreposição, como demonstrado na Tabela 7.

A capacidade de lidar com múltiplas instâncias de forma eficiente e precisa é crucial para a aplicação prática do modelo em ambientes reais, onde a sobreposição de objetos é comum.

Tabela 7 – Volume real medido, o RPE de cada variação e o MAPE estimado a partir de instâncias múltiplas de resíduos sólidos

Tipo	**Volume Real (L)**	**RPE 1**	**RPE 2**	**MAPE**
Resíduo Sólido Múltiplo 1	2.0744	1.18%	3.93%	2.55%
Resíduo Sólido Múltiplo 2	1.4299	1.72%	4.72%	3.22%
Resíduo Sólido Múltiplo 3	2.9652	1.53%	7.77%	4.65%
Resíduo Sólido Múltiplo 4	2.2495	1.28%	10.84%	6.06%
Resíduo Sólido Múltiplo 5	3.6752	7.67%	7.90%	7.78%
Resíduo Sólido Múltiplo 6	1.9745	7.60%	11.34%	9.47%
Resíduo Sólido Múltiplo 7	3.1460	10.13%	15.36%	12.74%
Resíduo Sólido Múltiplo 8	1.9555	17.45%	10.83%	14.14%
Resíduo Sólido Múltiplo 9	1.8683	11.90%	17%	14.45%
Resíduo Sólido Múltiplo 10	2.8377	17.23%	17.32%	17.28%

Fonte: Elaborado pelo autor

Na análise de múltiplos objetos, foram obtidos resultados muito próximos daqueles obtidos ao analisar apenas um único objeto. Nossa MAPE foi de 9.23%, com valores máximos de 17.28% e mínimos de 2.55%. Em comparação, o artigo de Graikos et al. (2020) obteve uma média de 77.41%, com um valor máximo de 97.28% e mínimo de 57.55%.

Como o conceito de testes em única ou múltiplas instâncias é introduzido no presente trabalho, não existem valores de comparação em relação ao coeficiente de determinação com outros artigos. Dessa forma, a comparação se mantém conforme apresentada na seção 5.3.1. No entanto, ao calcular o valor do R^2 com os dados dispostos na Tabela 7, foi obtido um valor de 87.64%. Embora seja consideravelmente menor em comparação com a única instância, ainda é um resultado satisfatório, considerando a complexidade envolvida na estimação de imagens com múltiplos objetos a serem analisados.

Esses resultados indicam que, em casos onde os objetos se sobrepõem, segmentar todas as instâncias juntas como um único item grande produziu resultados superiores. Embora essa abordagem possa dificultar a determinação precisa do volume individual de cada objeto, os ganhos em precisão geral superam essa limitação. Ao tratar múltiplas

instâncias como uma única entidade, foi possível minimizar os erros de sobreposição e obter uma estimativa de volume mais consistente e precisa.

5.3.3 Comparativo - Única x Múltiplas Instâncias

A comparação entre os resultados para objetos individuais e múltiplos revela nuances importantes na eficácia do modelo de estimativa de volume proposto. Para os objetos individuais, observamos uma alta precisão e consistência na estimativa de volume, caracterizada por erros que variam de baixos a moderados. Essa variação nos resultados pode ser atribuída às diferentes formas e tamanhos dos resíduos, além de variáveis como distância e outras condições gerais da imagem, que têm um impacto direto no resultado final.

Já na análise de múltiplos objetos, mesmo enfrentando o desafio da sobreposição, o modelo demonstrou uma capacidade robusta em estimar volumes. A abordagem de segmentação dos resíduos sobrepostos como um único item mostrou-se eficaz em mitigar erros decorrentes dessa complexidade adicional.

Os resultados indicam que, embora a precisão possa variar ligeiramente entre objetos individuais e múltiplos, o modelo mantém uma performance consistente e precisa em ambos os cenários. Isso sugere que a metodologia adotada é adequada para aplicações práticas na estimativa de volume de resíduos sólidos, proporcionando resultados confiáveis independentemente da complexidade da cena.

5.4 Considerações Finais

O Capítulo 5 apresentou os resultados do sistema proposto para o processo de estimação de volume de resíduos sólidos urbanos. Os resultados obtidos para objetos individuais e múltiplos demonstram que a integração de modelos, técnicas e algoritmos no sistema alcançou resultados promissores. Apesar das limitações na quantidade de dados disponíveis, o sistema mostrou-se eficaz na tarefa de estimação volumétrica desses itens.

É evidente que há espaço para melhorias significativas visando aprimorar a eficiência e a otimização geral do sistema. A análise dos resultados destacou áreas específicas que podem ser aprimoradas, como a automatização de variáveis críticas para a precisão da estimação, por exemplo, a determinação precisa da distância entre o objeto e a câmera. Tais melhorias não apenas têm o potencial de aumentar a precisão das estimativas, mas também de expandir a aplicabilidade do sistema em ambientes mais complexos e dinâmicos, permitindo uma adaptação mais eficiente a variações nas condições de medição, garantindo resultados mais confiáveis e consistentes na estimativa volumétrica de resíduos sólidos urbanos.

Portanto, os resultados obtidos fornecem uma base sólida para futuros desenvolvimentos e refinamentos do sistema, visando atender melhor às demandas práticas de gestão de resíduos sólidos urbanos através de técnicas avançadas de visão computacional e aprendizado de máquina.

6 Conclusões

O presente trabalho abordou a problemática da gestão de resíduos sólidos urbanos com foco na estimativa volumétrica a partir de imagens de visualização única, utilizando técnicas de visão computacional e aprendizado profundo. Os resultados obtidos demonstraram que os modelos propostos são promissores, apresentando altos níveis de precisão na segmentação de imagens e na estimação volumétrica de resíduos sólidos.

Utilizando as métricas MAPE e coeficiente de determinação, foi possível avaliar o desempenho geral do sistema proposto. Os resultados mostraram uma média de 8,60% para resíduos únicos e 9,23% para resíduos múltiplos, resultando em uma média geral de 8,91% para MAPE. Já para o coeficiente de determinação, foi obtido o valor de 95.11% para única instância de 87.64% para múltiplas. Esses resultados são satisfatórios, especialmente quando comparados com estudos na literatura que utilizam técnicas semelhantes.

A análise dos resultados revelou que a abordagem de segmentar múltiplos objetos como uma única entidade pode mitigar erros de sobreposição, proporcionando uma estimativa de volume mais consistente e precisa. Essa característica é crucial para a aplicação prática em ambientes urbanos, onde a densidade e a variabilidade dos resíduos podem influenciar diretamente na eficiência da coleta e do processamento.

O trabalho conseguiu atingir seu objetivo principal ao desenvolver um *framework* robusto e adaptável para a estimativa volumétrica de resíduos sólidos urbanos. É esperado que a precisão na detecção e estimativa do volume de resíduos identificados pelo modelo possa contribuir para a otimização das operações de coleta, transporte e processamento, resultando em um sistema de gestão de resíduos mais eficiente e sustentável.

A metodologia desenvolvida não apenas se mostrou eficaz no contexto estudado, mas também oferece potencial para ser aplicada em diferentes cenários e condições, mediante adaptações específicas. A utilização de técnicas de aprendizado profundo mostrou-se vantajosa, permitindo que o sistema se ajuste automaticamente a novas informações e condições ambientais, melhorando continuamente seu desempenho.

Entretanto, algumas limitações foram identificadas, como a necessidade de maior diversidade no conjunto de dados utilizado para treinamento e a adaptação dos modelos para diferentes cenários urbanos. Além disso, são necessárias melhorias a nível de código, como a otimização do uso de GPU para acelerar o processo de treino e facilitar testes com outros hiperparâmetros. Futuras pesquisas poderão focar na ampliação do *dataset* e na implementação de técnicas de aprendizado por transferência para melhorar ainda mais o desempenho dos modelos.

Em resumo, a dissertação atingiu seus objetivos ao propor uma solução inovadora e eficiente para a estimativa volumétrica de resíduos sólidos urbanos, contribuindo para a melhoria da gestão ambiental e para o desenvolvimento de cidades mais sustentáveis. A continuidade deste trabalho poderá proporcionar avanços significativos na área, promovendo a integração de tecnologias de ponta na gestão de resíduos e fortalecendo a sustentabilidade urbana.

6.1 Trabalhos Futuros

Para dar continuidade à pesquisa desenvolvida, são apresentados a seguir potenciais direcionamentos para desenvolvimentos futuros:

- Adicionar mais imagens reais e representativas do problema no conjunto de dados *WasteInsight*;
- Analisar a possibilidade de utilizar inteligências artificiais para criar imagens a serem adicionadas ao conjunto de dados;
- Otimizar todo o processo de treinamento do modelo de segmentação, tornando-o mais dinâmico e com maior capacidade de adicionar novas imagens e mais épocas;
- Adicionar suporte a GPU para melhorar a eficiência do processo de treinamento dos modelos;
- Criar um *Dataset* para o treino do modelo de estimativa de profundidade, que seja mais representativo em relação ao problema apresentado;
- Aplicar e analisar o funcionamento do sensor LiDAR para medir com precisão e de maneira dinâmica a distância do item analisado;
- Desenvolver uma forma mais precisa de estimar o volume real dos objetos irregulares analisados;
- Aplicar a estratégia proposta para teste em ambientes representativos do problema a ser solucionado;

6.2 Trabalhos Publicados

Publicações realizadas relativas à pesquisa durante o mestrado:

AZANCORT NETO, J. L. et al. Advanced single-view image-based framework for volume estimation in urban solid waste management. In: Anais do XV Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais. Porto Alegre, RS,

Brasil: SBC, 2024. p. 111–120. Disponível em: <https://sol.sbc.org.br/index.php/wcama/article/view/29423>.

AZANCORT NETO, J. L. et al. Wasteinsight: Conjunto de dados para detecção e estimativa de volume de resíduos sólidos urbanos. In: Anais do XV Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais (WCAMA 2024). Sociedade Brasileira de Computação - SBC, 2024. (WCAMA 2024). Disponível em: <http://dx.doi.org/10.5753/wcama.2024.2956>.

6.3 Dificuldades Encontradas

No desenvolvimento do sistema para a estimação volumétrica de resíduos sólidos urbanos baseado em visão computacional, foram enfrentadas várias dificuldades que impactaram diretamente o andamento e os resultados da pesquisa. Essas dificuldades foram diversas, abrangendo desde problemas técnicos até limitações relacionadas aos dados utilizados. A seguir, detalham-se as principais dificuldades encontradas:

- A variabilidade dos resíduos sólidos em termos de cor, formato, textura e tamanho representa um grande desafio, especialmente no contexto do descarte de lixo no Brasil. A falta de homogeneidade dos resíduos dificulta a criação de um modelo generalista e eficiente, pois os algoritmos de segmentação e detecção precisam lidar com uma vasta diversidade de características;
- A falta de dados de qualidade e representativos é uma das principais dificuldades enfrentadas. Dados insuficientes ou mal anotados prejudicam o treinamento dos modelos, resultando em estimativas menos precisas. Esse problema é exacerbado pela necessidade de grandes volumes de dados para o treinamento de redes neurais profundas. Mesmo com a criação do conjunto *WasteInsight*, torna-se evidente a necessidade de um volume maior de dados e de mais variações nos mesmos;
- Em cenários onde há múltiplos resíduos sobrepostos, a segmentação precisa torna-se mais complexa. A sobreposição de resíduos pode levar a falhas na segmentação e, consequentemente, na estimativa de profundidade e cálculo do volume. Para mitigar esses erros, foi adotada a abordagem de segmentar os resíduos como um único item nos casos onde não era possível distinguir claramente as bordas de um resíduo em relação ao outro;
- A necessidade de processamento intensivo, principalmente para o treinamento dos modelos de segmentação e estimativa de profundidade, exige recursos computacionais significativos. A falta de suporte adequado para GPU foi um limitador, impactando a eficiência do treinamento e a capacidade de adicionar novas imagens e mais épocas de treinamento;

- Estimar o volume de objetos com formas irregulares é particularmente desafiador. A precisão da estimativa depende da capacidade do modelo de interpretar corretamente as variações na forma e no tamanho dos resíduos, dada a complexidade da sua natureza irregular;
- Estimar o volume real de um objeto para compará-lo com o volume estimado pelo sistema também é um desafio. A variabilidade em termos de tipos e tamanhos de resíduos, além da natureza irregular de suas formas, torna extremamente complexo encontrar uma abordagem eficiente e generalista que lide adequadamente com essas diversas variações, impactando diretamente os resultados do sistema;
- Durante o desenvolvimento do sistema, foi necessário inserir manualmente os valores de distância da câmera para os objetos, o que introduziu uma margem de erro significativa.

Referências

ABREMA. *Índice de Sustentabilidade da Limpeza Urbana (ISLU)*. 2023. Disponível em: <https://www.abrema.org.br/islu/>. Citado 2 vezes nas páginas 7 e 8.

ALMAGHRABI, R. et al. A novel method for measuring nutrition intake based on food image. In: *2012 IEEE International Instrumentation and Measurement Technology Conference Proceedings*. [S.l.: s.n.], 2012. p. 366–370. Citado na página 32.

APARNA, H. et al. Iot assisted waste collection and management system using qr codes. In: IEEE. *2021 International Conference on Computer Communication and Informatics (ICCCI)*. [S.l.], 2021. p. 1–4. Citado na página 10.

ARBELáEZ-ESTRADA, J. C. et al. A systematic literature review of waste identification in automatic separation systems. *Recycling*, v. 8, n. 6, 2023. ISSN 2313-4321. Disponível em: <https://www.mdpi.com/2313-4321/8/6/86>. Citado na página 1.

AZANCORT NETO, J. L. et al. Artificial intelligence implemented to recognize patterns of sustainable areas by evaluating the database of socioenvironmental safety restrictions. *Research, Society and Development*, v. 10, n. 10, p. e212101018841–e212101018841, 2021. Citado 2 vezes nas páginas 1 e 9.

AZANCORT NETO, J. L. et al. Advanced single-view image-based framework for volume estimation in urban solid waste management. In: *Anais do XV Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais*. Porto Alegre, RS, Brasil: SBC, 2024. p. 111–120. Disponível em: <https://sol.sbc.org.br/index.php/wcama/article/view/29423>. Citado 3 vezes nas páginas 1, 2 e 47.

AZANCORT NETO, J. L. et al. Wasteinsight: Conjunto de dados para detecção e estimativa de volume de resíduos sólidos urbanos. In: *Anais do XV Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais (WCAMA 2024)*. Sociedade Brasileira de Computação - SBC, 2024. (WCAMA 2024). Disponível em: <http://dx.doi.org/10.5753/wcama.2024.2956>. Citado 5 vezes nas páginas 14, 48, 49, 51 e 56.

AZARAFZA, M.; KOÇKAR, M. K.; FARAMARZI, L. Spacing and block volume estimation in discontinuous rock masses using image processing technique: a case study. *Environmental Earth Sciences*, Springer, v. 80, n. 14, p. 471, 2021. Citado na página 38.

BANDI, N. et al. Image-based volume estimation using stereo vision. In: *2020 IEEE 18th International Symposium on Intelligent Systems and Informatics (SISY)*. [S.l.: s.n.], 2020. p. 000055–000060. Citado 6 vezes nas páginas 32, 33, 41, 44, 61 e 62.

BANO, A.; DIN, I. U.; AL-HUQAIL, A. A. Aiot-based smart bin for real-time monitoring and management of solid waste. *Scientific Programming*, Wiley Online Library, v. 2020, n. 1, p. 6613263, 2020. Citado na página 40.

BELLINI, V. et al. Reflective internet of things middleware-enabled a predictive real-time waste monitoring system. In: _____. *Web Engineering*. Springer International Publishing, 2018. p. 375–383. ISBN 9783319916620. Disponível em: <http://dx.doi.org/10.1007/978-3-319-91662-0_31>. Citado na página 10.

BRITO, D. et al. Manejo de resíduos sólidos e de águas pluviais: O (des)controle social em belém, pará. *Revista Eletrônica de Gestão e Tecnologias Ambientais*, v. 8, p. 103, 12 2020. Citado na página 3.

CHEN, P.-H. et al. Mvsnet++: Learning depth-based attention pyramid features for multi-view stereo. *IEEE Transactions on Image Processing*, v. 29, p. 7261–7273, 2020. Citado na página 45.

CHEN, W. et al. Single-image depth perception in the wild. *Advances in neural information processing systems*, v. 29, 2016. Citado na página 30.

CHEN, X. et al. Chinesefoodnet: A large-scale image dataset for chinese food recognition. *arXiv preprint arXiv:1705.02743*, 2017. Citado na página 39.

CHICCO, D.; WARRENS, M. J.; JURMAN, G. The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *PeerJ Computer Science*, PeerJ, v. 7, p. e623, jul. 2021. ISSN 2376-5992. Disponível em: <http://dx.doi.org/10.7717/peerj-cs.623>. Citado 2 vezes nas páginas 36 e 37.

CIOCCA, G.; NAPOLETANO, P.; SCHETTINI, R. Food recognition: a new dataset, experiments, and results. *IEEE journal of biomedical and health informatics*, IEEE, v. 21, n. 3, p. 588–598, 2016. Citado na página 39.

CIOCCA, G.; NAPOLETANO, P.; SCHETTINI, R. Food recognition: A new dataset, experiments, and results. *IEEE Journal of Biomedical and Health Informatics*, Institute of Electrical and Electronics Engineers (IEEE), v. 21, n. 3, p. 588–598, maio 2017. ISSN 2168-2208. Disponível em: <http://dx.doi.org/10.1109/jbhi.2016.2636441>. Citado na página 42.

DALAI, R.; DALAI, N.; SENAPATI, K. K. An accurate volume estimation on single view object images by deep learning based depth map analysis and 3d reconstruction. *Multimedia Tools and Applications*, Springer Science and Business Media LLC, v. 82, n. 18, p. 28235–28258, fev. 2023. ISSN 1573-7721. Disponível em: <http://dx.doi.org/10.1007/s11042-023-14615-7>. Citado 9 vezes nas páginas 31, 32, 35, 36, 42, 44, 55, 61 e 62.

DAMEN, D. et al. The epic-kitchens dataset: Collection, challenges and baselines. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 43, n. 11, p. 4125–4141, 2021. Citado 4 vezes nas páginas 39, 42, 49 e 53.

DEHAIS, J. et al. Two-view 3d reconstruction for food volume estimation. *IEEE Transactions on Multimedia*, v. 19, n. 5, p. 1090–1099, 2017. Citado 5 vezes nas páginas 32, 35, 41, 43 e 55.

DHILLON, A.; VERMA, G. K. Convolutional neural network: a review of models, methodologies and applications to object detection. *Progress in Artificial Intelligence*,

Springer Science and Business Media LLC, v. 9, n. 2, p. 85–112, dez. 2019. ISSN 2192-6360. Disponível em: <http://dx.doi.org/10.1007/s13748-019-00203-0>. Citado na página 18.

DIALOGO, G. G. Fish species detection application (fisda) in leyte gulf using convolutional neural network. *Proceedings of Engineering and Technology Innovation*, Taiwan Association of Engineering and Technology Innovation, v. 19, p. 16, 2021. Citado na página 18.

DONG, S.; WANG, P.; ABBAS, K. A survey on deep learning and its applications. *Computer Science Review*, Elsevier, v. 40, p. 100379, 2021. Citado na página 11.

DUTTA, A.; GUPTA, A.; ZISSERMANN, A. *VGG image annotator (VIA)*. 2016. Citado na página 49.

ECK, N. J. V.; WALTMAN, L. Vosviewer manual. *Manual for VOSviewer version*, v. 1, n. 0, 2011. Citado na página 40.

EDELSBRUNNER, H.; HARER, J. L. *Computational topology: an introduction*. [S.l.]: American Mathematical Society, 2022. Citado na página 54.

ESTANQUEIRO JOSé DINIS SILVESTRE, J. d. B. B.; PINHEIRO, M. D. Environmental life cycle assessment of coarse natural and recycled aggregates for concrete. *European Journal of Environmental and Civil Engineering*, Taylor & Francis, v. 22, n. 4, p. 429–449, 2018. Citado na página 8.

ESTEVA, A. et al. Deep learning-enabled medical computer vision. *NPJ digital medicine*, Nature Publishing Group UK London, v. 4, n. 1, p. 5, 2021. Citado na página 13.

FATANIYA, B. et al. Implementation of iot based waste segregation and collection system. *International Journal of Electronics and Telecommunications*, Polish Academy of Sciences Chancellery, p. 579–584, jul. 2019. ISSN 2300-1933. Disponível em: <http://dx.doi.org/10.24425/ijet.2019.129816>. Citado na página 10.

FERRER, J.; ALBA, E. Bin-ct: Urban waste collection based on predicting the container fill level. *Biosystems*, Elsevier BV, v. 186, p. 103962, dez. 2019. ISSN 0303-2647. Disponível em: <http://dx.doi.org/10.1016/j.biosystems.2019.04.006>. Citado na página 10.

FERRONATO, N.; TORRETTA, V. Waste mismanagement in developing countries: A review of global issues. *International journal of environmental research and public health*, MDPI, v. 16, n. 6, p. 1060, 2019. Citado na página 8.

GALLARDO, A. et al. Methodology to design a municipal solid waste pre-collection system. a case study. *Waste Management*, Elsevier BV, v. 36, p. 1–11, fev. 2015. ISSN 0956-053X. Disponível em: <http://dx.doi.org/10.1016/j.wasman.2014.11.008>. Citado na página 8.

GATTI, A.; BARBIERATO, E.; POZZI, A. Toward greener smart cities: A critical review of classic and machine-learning-based algorithms for smart bin collection. *Electronics*, MDPI, v. 13, n. 5, p. 836, 2024. Citado na página 40.

GHOSH, A. et al. Fundamental concepts of convolutional neural network. *Recent trends and advances in artificial intelligence and Internet of Things*, Springer, p. 519–567, 2020. Citado 6 vezes nas páginas 20, 21, 22, 23, 25 e 27.

GIRSHICK, R. Fast r-cnn. In: *Proceedings of the IEEE international conference on computer vision*. [S.l.: s.n.], 2015. p. 1440–1448. Citado 2 vezes nas páginas 16 e 50.

GODARD, C.; AODHA, O. M.; BROSTOW, G. J. Unsupervised monocular depth estimation with left-right consistency. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 270–279. Citado 2 vezes nas páginas 30 e 45.

GODARD, C. et al. Digging into self-supervised monocular depth estimation. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 2019. Citado 5 vezes nas páginas 31, 33, 48, 52 e 53.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep learning*. [S.l.]: MIT press, 2016. Citado na página 16.

GRAIKOS, A. et al. Single image-based food volume estimation using monocular depth-prediction networks. In: SPRINGER. *Universal Access in Human-Computer Interaction. Applications and Practice: 14th International Conference, UAHCI 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part II 22*. [S.l.], 2020. p. 532–543. Citado 15 vezes nas páginas 15, 32, 33, 35, 42, 43, 44, 48, 49, 51, 53, 54, 61, 62 e 63.

GRECHINSKI, P. Lixo no mar. *Revista Mosaicos: Estudos em Governança, Sustentabilidade e Inovação*, Revista Mosaicos, v. 2, n. 1, p. 30–43, dez. 2020. ISSN 2674-8258. Disponível em: <http://dx.doi.org/10.37032/remos.v2i1.31>. Citado na página 6.

GU, J. et al. Recent advances in convolutional neural networks. *Pattern recognition*, Elsevier, v. 77, p. 354–377, 2018. Citado na página 24.

GUERRERO, L. A.; MAAS, G.; HOGLAND, W. Solid waste management challenges for cities in developing countries. *Waste Management*, v. 33, n. 1, p. 220–232, 2013. ISSN 0956-053X. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0956053X12004205>. Citado na página 1.

HAFIZ, A. M.; BHAT, G. M. A survey on instance segmentation: state of the art. *International journal of multimedia information retrieval*, Springer, v. 9, n. 3, p. 171–189, 2020. Citado na página 15.

HASSANNEJAD, H. et al. A new approach to image-based estimation of food volume. *Algorithms*, MDPI AG, v. 10, n. 2, p. 66, jun. 2017. ISSN 1999-4893. Disponível em: <http://dx.doi.org/10.3390/a10020066>. Citado 3 vezes nas páginas 32, 33 e 41.

HAYWOOD, L. K. et al. Waste disposal practices in low-income settlements of south africa. *International Journal of Environmental Research and Public Health*, MDPI AG, v. 18, n. 15, p. 8176, ago. 2021. ISSN 1660-4601. Disponível em: <http://dx.doi.org/10.3390/ijerph18158176>. Citado 2 vezes nas páginas 7 e 9.

HE, K. et al. Mask r-cnn. In: *Proceedings of the IEEE international conference on computer vision*. [S.l.: s.n.], 2017. p. 2961–2969. Citado 2 vezes nas páginas 16 e 50.

HE, K. et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 770–778. Citado 2 vezes nas páginas 28 e 29.

HE, L. et al. Sosd-net: Joint semantic object segmentation and depth estimation from monocular images. *Neurocomputing*, Elsevier, v. 440, p. 251–263, 2021. Citado na página 45.

HUANG, G. et al. Densely connected convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 4700–4708. Citado na página 16.

HUYNH, T. T.; TONTHAT, L.; DAO, S. V. A vision-based method to estimate volume and mass of fruit/vegetable: Case study of sweet potato. *International Journal of Food Properties*, Taylor & Francis, v. 25, n. 1, p. 717–732, 2022. Citado na página 40.

ISABONA, J. et al. Development of a multilayer perceptron neural network for optimal predictive modeling in urban microcellular radio environments. *Applied Sciences*, MDPI, v. 12, n. 11, p. 5713, 2022. Citado na página 22.

JADHAV, T.; SINGH, K.; ABHYANKAR, A. Volumetric estimation using 3d reconstruction method for grading of fruits. *Multimedia Tools and Applications*, Springer, v. 78, p. 1613–1634, 2019. Citado na página 32.

JAIN, R. et al. Convolutional neural network based alzheimer's disease classification from magnetic resonance brain images. *Cognitive Systems Research*, Elsevier, v. 57, p. 147–159, 2019. Citado na página 14.

JANIESCH, C.; ZSCHECH, P.; HEINRICH, K. Machine learning and deep learning. *Electronic Markets*, Springer, v. 31, n. 3, p. 685–695, 2021. Citado na página 11.

JESCA, M.; JUNIOR, M. Practices regarding disposal of soiled diapers among women of child bearing age in poor resource urban setting. *J. Nurs. Health Sci*, Citeseer, v. 4, n. 4, p. 63–67, 2015. Citado na página 8.

KAKANI, V. et al. A critical review on computer vision and artificial intelligence in food industry. *Journal of Agriculture and Food Research*, Elsevier, v. 2, p. 100033, 2020. Citado na página 13.

KAWANO, Y.; YANAI, K. Automatic expansion of a food image dataset leveraging existing categories with domain adaptation. In: SPRINGER. *Computer Vision-ECCV 2014 Workshops: Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part III 13*. [S.l.], 2015. p. 3–17. Citado na página 39.

KETKAR, N.; MOOLAYIL, J. Convolutional neural networks. In: _____. *Deep Learning with Python*. Apress, 2021. p. 197–242. ISBN 9781484253649. Disponível em: <http://dx.doi.org/10.1007/978-1-4842-5364-9_6>. Citado 4 vezes nas páginas 17, 18, 19 e 21.

KHAN, A. et al. A survey of the recent architectures of deep convolutional neural networks. *Artificial intelligence review*, Springer, v. 53, p. 5455–5516, 2020. Citado 2 vezes nas páginas 20 e 24.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, v. 25, 2012. Citado 2 vezes nas páginas 24 e 25.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, AcM New York, NY, USA, v. 60, n. 6, p. 84–90, 2017. Citado 2 vezes nas páginas 16 e 34.

LAWSON, N. et al. Recycling construction and demolition wastes–a uk perspective. *Environmental Management and Health*, MCB UP Ltd, v. 12, n. 2, p. 146–157, 2001. Citado na página 8.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *nature*, Nature Publishing Group UK London, v. 521, n. 7553, p. 436–444, 2015. Citado na página 16.

LECUN, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, Ieee, v. 86, n. 11, p. 2278–2324, 1998. Citado 3 vezes nas páginas 16, 23 e 24.

LI, H.; HAN, T. Deepvol: Deep fruit volume estimation. In: SPRINGER. *Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III 27*. [S.l.], 2018. p. 331–341. Citado na página 30.

LI, N. et al. A progress review on solid-state lidar and nanophotonics-based lidar sensors. *Laser & Photonics Reviews*, Wiley Online Library, v. 16, n. 11, p. 2100511, 2022. Citado na página 29.

LI, Y. et al. Fully convolutional instance-aware semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 2359–2367. Citado na página 16.

LI, Z. et al. A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*, v. 33, n. 12, p. 6999–7019, 2022. Citado 2 vezes nas páginas 17 e 18.

LIANG, Y.; LI, J. Deep learning-based food calorie estimation method in dietary assessment. *CoRR*, abs/1706.04062, 2017. Disponível em: <http://arxiv.org/abs/1706.04062>. Citado na página 32.

LIAO, J. et al. Adaptive depth estimation for pyramid multi-view stereo. *Computers amp; Graphics*, Elsevier BV, v. 97, p. 268–278, jun. 2021. ISSN 0097-8493. Disponível em: <http://dx.doi.org/10.1016/j.cag.2021.04.016>. Citado 2 vezes nas páginas 31 e 45.

LIN, T.-Y. et al. Feature pyramid networks for object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 2117–2125. Citado na página 16.

LIN, T.-Y. et al. Microsoft coco: Common objects in context. In: SPRINGER. *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*. [S.l.], 2014. p. 740–755. Citado 3 vezes nas páginas 16, 34 e 49.

LIU, F.; SHEN, C.; LIN, G. Deep convolutional neural fields for depth estimation from a single image. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2015. Citado na página 30.

LIU, W. et al. A survey of deep neural network architectures and their applications. *Neurocomputing*, Elsevier, v. 234, p. 11–26, 2017. Citado na página 16.

LO, F. P.-W. et al. Food volume estimation based on deep learning view synthesis from a single depth map. *Nutrients*, MDPI, v. 10, n. 12, p. 2005, 2018. Citado 5 vezes nas páginas 29, 32, 41, 42 e 44.

LO, F. P. W. et al. Image-based food classification and volume estimation for dietary assessment: A review. *IEEE journal of biomedical and health informatics*, IEEE, v. 24, n. 7, p. 1926–1939, 2020. Citado 2 vezes nas páginas 39 e 43.

LU, W.; CHEN, J. Computer vision for solid waste sorting: A critical review of academic research. *Waste Management*, Elsevier, v. 142, p. 29–43, 2022. Citado na página 2.

LU, Y. et al. A multi-task learning approach for meal assessment. In: *Proceedings of the Joint Workshop on Multimedia for Cooking and Eating Activities and Multimedia Assisted Dietary Management*. New York, NY, USA: Association for Computing Machinery, 2018. (CEA/MADiMa '18), p. 46–52. ISBN 9781450365376. Disponível em: <https://doi.org/10.1145/3230519.3230593>. Citado 2 vezes nas páginas 29 e 45.

LUNETTA, A. D.; GUERRA, R. Metodologia da pesquisa científica e acadêmica. *Revista OWL (OWL Journal)-Revista Interdisciplinar de Ensino e Educação*, v. 1, n. 2, p. 149–159, 2023. Citado na página 38.

MAHMOOD, I.; ZUBAIRI, J. A. Efficient waste transportation and recycling: Enabling technologies for smart cities using the internet of things. *IEEE Electrification Magazine*, v. 7, n. 3, p. 33–43, 2019. Citado na página 10.

MATHEW, A.; AMUDHA, P.; SIVAKUMARI, S. Deep learning techniques: an overview. *Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2020*, Springer, p. 599–608, 2021. Citado na página 11.

MATSUO, Y. et al. Deep learning, reinforcement learning, and world models. *Neural Networks*, Elsevier, v. 152, p. 267–275, 2022. Citado na página 11.

MERTAN, A.; DUFF, D. J.; UNAL, G. Single image depth estimation: An overview. *Digital Signal Processing*, Elsevier, v. 123, p. 103441, 2022. Citado na página 30.

MEYERS, A. et al. Im2calories: Towards an automated mobile vision food diary. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 2015. Citado na página 30.

MINAEE, S. et al. Image segmentation using deep learning: A survey. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 44, n. 7, p. 3523–3542, 2021. Citado na página 15.

MING, Y. et al. Deep learning for monocular depth estimation: A review. *Neurocomputing*, Elsevier, v. 438, p. 14–33, 2021. Citado na página 30.

MOYA, D. et al. Waste-to-energy technologies: an opportunity of energy recovery from municipal solid waste, using quito-ecuador as case study. *Energy Procedia*, Elsevier, v. 134, p. 327–336, 2017. Citado na página 6.

MYTTENAERE, A. D. et al. Mean absolute percentage error for regression models. *Neurocomputing*, Elsevier, v. 192, p. 38–48, 2016. Citado na página 36.

NANDA, S.; BERRUTI, F. Municipal solid waste management and landfilling technologies: a review. *Environmental chemistry letters*, Springer, v. 19, n. 2, p. 1433–1456, 2021. Citado na página 5.

NANDA, S.; BERRUTI, F. Thermochemical conversion of plastic waste to fuels: a review. *Environmental Chemistry Letters*, Springer, v. 19, n. 1, p. 123–148, 2021. Citado na página 6.

OKINDA, C. et al. Egg volume estimation based on image processing and computer vision. *Journal of Food Engineering*, Elsevier BV, v. 283, p. 110041, out. 2020. ISSN 0260-8774. Disponível em: <http://dx.doi.org/10.1016/j.jfoodeng.2020.110041>. Citado 3 vezes nas páginas 29, 39 e 45.

OKPARA, D. A.; KHARLAMOVA, M.; GRACHEV, V. Proliferation of household waste irregular dumpsites in niger delta region (nigeria): unsustainable public health monitoring and future restitution. *Sustainable Environment Research*, Springer Science and Business Media LLC, v. 31, n. 1, jan. 2021. ISSN 2468-2039. Disponível em: <http://dx.doi.org/10.1186/s42834-020-00077-1>. Citado na página 6.

PADILLA, R.; NETTO, S. L.; SILVA, E. A. B. da. A survey on performance metrics for object-detection algorithms. In: *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*. [S.l.: s.n.], 2020. p. 237–242. Citado 2 vezes nas páginas 33 e 34.

PADILLA, R. et al. A comparative analysis of object detection metrics with a companion open-source toolkit. *Electronics*, MDPI, v. 10, n. 3, p. 279, 2021. Citado na página 34.

PARDINI, K. et al. Iot-based solid waste management solutions: a survey. *Journal of Sensor and Actuator Networks*, MDPI, v. 8, n. 1, p. 5, 2019. Citado na página 10.

PEREIRA, A. S. et al. *Metodologia da pesquisa científica*. Universidade Federal de Santa Maria, 2018. Disponível em: <http://repositorio.ufsm.br/handle/1/15824>. Citado na página 38.

POPLY, P.; JOTHI, J. A. A. Refined image segmentation for calorie estimation of multiple-dish food items. In: IEEE. *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*. [S.l.], 2021. p. 682–687. Citado na página 39.

PRIYA, K. S. et al. Impact of ammonia nitrogen on cod removal efficiency in anaerobic hybrid membrane bioreactor treating synthetic leachate. *International Journal of Environmental Research*, Springer, v. 13, p. 59–65, 2019. Citado na página 9.

PURI, M. et al. Recognition and volume estimation of food intake using a mobile device. In: *2009 Workshop on Applications of Computer Vision (WACV)*. [S.l.: s.n.], 2009. p. 1–8. Citado na página 32.

QI, X. et al. Geonet: Geometric neural network for joint depth and surface normal estimation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2018. p. 283–291. Citado na página 30.

RABBANI, A. et al. Image-based estimation of the left ventricular cavity volume using deep learning and gaussian process with cardio-mechanical applications. *Computerized Medical Imaging and Graphics*, Elsevier BV, v. 106, p. 102203, jun. 2023. ISSN 0895-6111. Disponível em: <http://dx.doi.org/10.1016/j.compmedimag.2023.102203>. Citado na página 38.

RAITOHARJU, J. Convolutional neural networks. In: _____. *Deep Learning for Robot Perception and Cognition*. Elsevier, 2022. p. 35–69. ISBN 9780323857871. Disponível em: <http://dx.doi.org/10.1016/B978-0-32-385787-1.00008-7>. Citado na página 19.

REN, S. et al. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 39, n. 6, p. 1137–1149, 2016. Citado na página 16.

ROBERTS, D. A.; YAIDA, S.; HANIN, B. *The principles of deep learning theory*. [S.l.]: Cambridge University Press Cambridge, MA, USA, 2022. v. 46. Citado 2 vezes nas páginas 11 e 12.

ROY, A. et al. Iot-based smart bin allocation and vehicle routing in solid waste management: A case study in south korea. *Computers & Industrial Engineering*, Elsevier, v. 171, p. 108457, 2022. Citado na página 40.

SAINATH, T. N. et al. Improvements to deep convolutional neural networks for LVCSR. *CoRR*, abs/1309.1501, 2013. Disponível em: <http://arxiv.org/abs/1309.1501>. Citado na página 22.

SEPADI, M. M. Unsafe management of soiled nappies in informal settlements and villages of south africa. *Cities & Health*, Taylor & Francis, v. 6, n. 2, p. 254–257, 2022. Citado na página 8.

SEROR, N.; HARELI, S.; PORTNOV, B. A. Evaluating the effect of vehicle impoundment policy on illegal construction and demolition waste dumping: Israel as a case study. *Waste management*, Elsevier, v. 34, n. 8, p. 1436–1445, 2014. Citado na página 8.

SHARMA, N.; SHARMA, R.; JINDAL, N. Machine learning and deep learning applications-a vision. *Global Transitions Proceedings*, Elsevier, v. 2, n. 1, p. 24–28, 2021. Citado na página 13.

SIDHU, N. et al. A collaborative application for assisting the management of household plastic waste through smart bins: a case of study in the philippines. *Sensors*, MDPI, v. 21, n. 13, p. 4534, 2021. Citado na página 10.

SILVA, A. S. et al. Capacitated waste collection problem solution using an open-source tool. *Computers*, MDPI, v. 12, n. 1, p. 15, 2023. Citado 2 vezes nas páginas 10 e 11.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. Citado na página 25.

SOH, Z. H. C. et al. Smart waste collection monitoring and alert system via iot. In: *2019 IEEE 9th Symposium on Computer Applications Industrial Electronics (ISCAIE)*. [S.l.: s.n.], 2019. p. 50–54. Citado na página 10.

SOUSA, A. S. de; OLIVEIRA, G. S. de; ALVES, L. H. A pesquisa bibliográfica: princípios e fundamentos. *Cadernos da FUCAMP*, v. 20, n. 43, 2021. Citado na página 39.

SRIVASTAVA, S. et al. Comparative analysis of deep learning image detection algorithms. *Journal of Big data*, Springer, v. 8, n. 1, p. 66, 2021. Citado 2 vezes nas páginas 14 e 15.

STEINBRENER, J. et al. Learning metric volume estimation of fruits and vegetables from short monocular video sequences. *Heliyon*, Elsevier, v. 9, n. 4, 2023. Citado 2 vezes nas páginas 32 e 55.

SU, Z. et al. View synthesis from multi-view rgb data using multilayered representation and volumetric estimation. *Virtual Reality & Intelligent Hardware*, Elsevier, v. 2, n. 1, p. 43–55, 2020. Citado 2 vezes nas páginas 16 e 44.

SULEMAN, Y.; DARKO, E.; AGYEMANG-DUAH, W. Solid waste disposal and community health implications in ghana: Evidence from sawaba, asokore mampong municipal assembly. *J Civil Environ Eng*, v. 5, n. 6, p. 202, 2015. Citado 2 vezes nas páginas 8 e 9.

SZEGEDY, C. et al. Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2015. p. 1–9. Citado 2 vezes nas páginas 27 e 28.

TAM, V. W. Comparing the implementation of concrete recycling in the australian and japanese construction industries. *Journal of Cleaner Production*, Elsevier BV, v. 17, n. 7, p. 688–702, maio 2009. ISSN 0959-6526. Disponível em: <http://dx.doi.org/10.1016/j.jclepro.2008.11.015>. Citado na página 8.

TAMMINA, S. Transfer learning using vgg-16 with deep convolutional neural network for classifying images. *International Journal of Scientific and Research Publications (IJSRP)*, International Journal of Scientific and Research Publications (IJSRP), v. 9, n. 10, p. p9420, out. 2019. ISSN 2250-3153. Disponível em: <http://dx.doi.org/10.29322/IJSRP.9.10.2019.p9420>. Citado na página 26.

TAMULY, S.; JYOTSNA, C.; AMUDHA, J. Deep learning model for image classification. In: SPRINGER. *Computational Vision and Bio-Inspired Computing: ICCVBIC 2019*. [S.l.], 2020. p. 312–320. Citado 2 vezes nas páginas 13 e 14.

TAşKıN, A.; DEMIR, N. Life cycle environmental and energy impact assessment of sustainable urban municipal solid waste collection and transportation strategies. *Sustainable Cities and Society*, Elsevier BV, v. 61, p. 102339, out. 2020. ISSN 2210-6707. Disponível em: <http://dx.doi.org/10.1016/j.scs.2020.102339>. Citado na página 6.

The World Bank. *Trends in solid waste management*. 2020. Disponível em: <https://datatopics.worldbank.org/what-a-waste/trends_in_solid_waste_management.html>. Citado na página 5.

THüRER, M. et al. Internet of things (iot) driven kanban system for reverse logistics: solid waste collection. *Journal of Intelligent Manufacturing*, Springer Science and Business Media LLC, v. 30, n. 7, p. 2621–2630, dez. 2016. ISSN 1572-8145. Disponível em: <http://dx.doi.org/10.1007/s10845-016-1278-y>. Citado na página 10.

UN-HABITAT. *Solid waste management in the world's cities: Water and sanitation in the world's cities 2010*. [S.l.]: Routledge, 2010. Citado na página 8.

Waste Atlas. *What a waste: an updated look into the future of solid waste management*. 2018. Disponível em: <https://www.worldbank.org/en/news/immersive-story/2018/09/20/what-a-waste-an-updated-look-into-the-future-of-solid-waste-management>. Citado na página 5.

WRIGHT, S. Correlation and causation. *Journal of agricultural research*, v. 20, n. 7, p. 557, 1921. Citado na página 36.

WU, H. et al. An innovative approach to managing demolition waste via gis (geographic information system): a case study in shenzhen city, china. *Journal of Cleaner Production*, Elsevier BV, v. 112, p. 494–503, jan. 2016. ISSN 0959-6526. Disponível em: <http://dx.doi.org/10.1016/j.jclepro.2015.08.096>. Citado na página 8.

XIE, J.; GIRSHICK, R.; FARHADI, A. Deep3d: Fully automatic 2d-to-3d video conversion with deep convolutional neural networks. In: SPRINGER. *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*. [S.l.], 2016. p. 842–857. Citado 3 vezes nas páginas 30, 31 e 45.

XU, C. et al. Image-based food volume estimation. In: *Proceedings of the 5th International Workshop on Multimedia for Cooking & Eating Activities*. New York, NY, USA: Association for Computing Machinery, 2013. (CEA '13), p. 75–80. ISBN 9781450323925. Disponível em: <https://doi.org/10.1145/2506023.2506037>. Citado na página 32.

XU, D. et al. Pad-net: Multi-tasks guided prediction-and-distillation network for simultaneous depth estimation and scene parsing. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2018. p. 675–684. Citado na página 30.

XU, S. et al. Computer vision techniques in construction: a critical review. *Archives of Computational Methods in Engineering*, Springer, v. 28, p. 3383–3397, 2021. Citado na página 13.

YANG, Z. et al. Human-mimetic estimation of food volume from a single-view rgb image using an ai system. *Electronics*, MDPI, v. 10, n. 13, p. 1556, 2021. Citado 4 vezes nas páginas 15, 29, 41 e 44.

YE, G. et al. Simulating effects of management measures on the improvement of the environmental performance of construction waste management. *Resources, conservation and recycling*, Elsevier, v. 62, p. 56–63, 2012. Citado na página 8.

YU, A. et al. Attention aware cost volume pyramid based multi-view stereo network for 3d reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, Elsevier BV, v. 175, p. 448–460, maio 2021. ISSN 0924-2716. Disponível em: <http://dx.doi.org/10.1016/j.isprsjprs.2021.03.010>. Citado na página 45.

ZHANG, H.; YANG, X. A survey on algorithms and performance metrics for object detection. *International Core Journal of Engineering*, Boya Century Publishing Limited, v. 9, n. 9, p. 133–145, 2023. Citado na página 35.

ZHANG, X. et al. Source separation, transportation, pretreatment, and valorization of municipal solid waste: a critical review. *Environment, Development and Sustainability*, Springer, p. 1–43, 2022. Citado na página 1.

ZHANG, Z. et al. Pattern-affinitive propagation across depth, surface normal and semantic segmentation. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. [S.l.: s.n.], 2019. p. 4106–4115. Citado na página 30.

ZHAO, C. et al. Monocular depth estimation based on deep learning: An overview. *Science China Technological Sciences*, Springer, v. 63, n. 9, p. 1612–1627, 2020. Citado na página 30.

ZHAO, X. et al. A review of convolutional neural networks in computer vision. *Artificial Intelligence Review*, Springer Science and Business Media LLC, v. 57, n. 4, mar. 2024. ISSN 1573-7462. Disponível em: <http://dx.doi.org/10.1007/s10462-024-10721-6>. Citado na página 22.

ZHOU, L.; ZHANG, L.; KONZ, N. Computer vision techniques in manufacturing. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, IEEE, v. 53, n. 1, p. 105–117, 2022. Citado na página 13.

ZHOU, T. et al. Unsupervised learning of depth and ego-motion from video. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2017. Citado 4 vezes nas páginas 31, 33, 52 e 54.

ZHU, Z. et al. Transfer learning in deep reinforcement learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 45, n. 11, p. 13344–13362, 2023. Citado na página 33.