UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

Aprendizagem profunda aplicada à separação de materiais recicláveis

Thais Hanashiro Moraes

Trabalho de Conclusão de Curso
MBA em Inteligência Artificial e Big Data

ICMC USP
SÃO CARLOS

UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

Aprendizagem profunda aplicada à separação de materiais recicláveis

Thais Hanashiro Moraes

USP - São Carlos

2022

Thais Hanashiro Moraes

Aprendizagem profunda aplicada à separação de materiais recicláveis

Trabalho de conclusão de curso apresentado ao Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, como parte dos requisitos para obtenção do título de Especialista em Inteligência Artificial e Big Data.

Área de concentração: Inteligência Artificial

Orientador: Prof. Dr. Alfredo Colenci Neto

USP - São Carlos

2022

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

M827a

Moraes, Thais Hanashiro
Aprendizagem profunda aplicada à separação de materiais recicláveis / Thais Hanashiro Moraes; orientador Alfredo Colenci Neto. -- São Carlos, 2022.
51 p.

Trabalho de conclusão de curso (MBA em Inteligência Artificial e Big Data) -- Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2022.

1. Redes Neurais. 2. Segmentação de imagens. 3. Reciclagem. I. Colenci Neto, Alfredo, orient. II. Título.

Bibliotecários responsáveis pela estrutura de catalogação da publicação de acordo com a AACR2:
Gláucia Maria Saia Cristianini - CRB - 8/4938
Juliana de Souza Moraes - CRB - 8/6176

RESUMO

MORAES, T. H. **Aprendizagem profunda aplicada à separação de materiais recicláveis**. 2022. 51 f. Trabalho de conclusão de curso (MBA em Inteligência Artificial e Big Data) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2022.

A gestão de resíduos sólidos urbanos (RSU) é um dos principais desafios que o mundo enfrenta atualmente (GUPTA et al., 2019). Anualmente são gerados 2.01 bilhões de toneladas de RSU, estima-se que esse número chegue a 3.40 bilhões até 2050. Desse total, ao menos 33% não são tratados de forma adequada, gerando impactos ambientais e sociais negativos (KAZA et al., 2018). O processo de reciclagem é essencial para a obtenção de uma economia circular, visando otimizar a etapa de triagem dos materiais recicláveis propõe-se utilização de uma rede neural profunda para a realização das tarefas de identificação e classificação necessárias. Para determinação da melhor arquitetura para a tarefa, inicialmente faz-se um levantamento bibliográfico, passando pela definição de redes neurais, até comentários sobre estruturas dos modelos estado da arte para detecção de objetos. Opta-se pela análise do impacto de alteração do backbone sobre a solução de segmentação de instância Mask R-CNN sobre a análise da desafiadora base ZeroWaste Dataset (BASHKIROVA, 2021). Tal dataset lançado em 2021, traz mais de 6 mil imagens anotadas com formato COCO, trazendo ao público uma base de domínio específico, voltada para a reciclagem de materiais em ambiente produtivo (quadros de filmagens feitas sobre uma esteira de separação de RSU recicláveis). Compara-se os resultados obtidos para a detecção de máscaras das instâncias anotadas para um modelo com backbone Resnet-50 e ResNext-101. Obtêm-se resultados a partir de 70% melhores para a precisão das máscaras com a backbone mais robusta.

Palavras-chave: aprendizagem profunda; materiais recicláveis; segmentação de instâncias; Mask R-CNN

ABSTRACT

MORAES, T. H. Deep learning applied to the separation of recyclable materials. 2022. 51 f. Trabalho de conclusão de curso (MBA em Inteligência Artificial e Big Data) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2022.

The management of municipal solid waste (MSW) is one of the main challenges our society currently faces (GUPTA et al., 2019). 2.01 billion tons of MSW are generated annually, it is estimated that this number will reach 3.40 billion by 2050. From this total, at least 33% are not adequately treated, generating negative environmental and social impacts (KAZA et al., 2018). The recycling process is a key component to reach a circular economy. Aiming to optimize the stage of sorting recyclable materials, it is proposed the use of a deep neural network to carry out the necessary detection and classification tasks. To set the best architecture for the task, initially an extense literature review is done, discussing topics from the definition of neural networks to comments on the structures of state-of-the-art models for object detection. We chose to analyze the impact of changing the backbone on the Mask R-CNN instance segmentation solution, when trained and tested on the challenging ZeroWaste Dataset base (BASHKIROVA, 2021). This dataset, launched in 2021, consists of more than 6,000 images annotated with COCO format, a public dataset regarding a specific domain base, focused on the recycling of materials in an industrial setting (frames obtained from a recycling paper facility's conveyor belt). The results of the segmentation task (masks) obtained from a model built on Resnet-50 are compared to the ones generated from a ResNext-101 backbone-based model. Results suggests a minimum improvement of 70% for the accuracy of the masks predictions with the most robust backbone network.

Keywords: deep learning; recyclable materials; instance segmentation; Mask R-CNN

SUMÁRIO

1 INTRODUÇÃO

1.1 Contextualização

A gestão de resíduos sólidos urbanos (RSU) é um dos principais desafios que o mundo enfrenta atualmente (GUPTA et al., 2019). Anualmente são gerados 2.01 bilhões de toneladas de RSU, estima-se que esse número chegue a 3.40 bilhões até 2050. Desse total, ao menos 33% não são tratados de forma adequada, gerando impactos ambientais e sociais negativos (poluição marinha, contaminação do solo e de lençóis freáticos, problemas de saúde pública, etc.) (KAZA et al., 2018). Gerir de forma eficiente os RSU é, portanto, de fundamental importância para construir-se uma sociedade sustentável.

Uma gestão eficiente dos RSU engloba a aplicação e otimização de diversas estratégias: redução, reuso, reciclagem, coleta e disposição final de resíduos orgânicos e inorgânicos. Este projeto tem como objetivo otimizar o processo de reciclagem, almejando facilitar a transição do modelo econômico-produtivo atual para um modelo de economia circular.

O processo de reciclagem poder ser dividido em quatro grandes etapas: separação e coleta seletiva, triagem dos materiais, comercialização e processamento industrial. A triagem dos materiais, em geral, é feita manualmente, sendo esse o gargalo da cadeia produtiva (PARREIRA, OLIVEIRA, LIMA, 2009), ou seja, o fator que limita a produtividade do processo de reciclagem. Almejando a otimização da etapa de triagem, propõe-se utilização de uma rede neural profunda para a realização das tarefas de identificação e classificação dos materiais recicláveis.

CNN’s *(convolutional neural network*) utilizam uma arquitetura bem adaptada para a tarefa de classificação de imagens, sendo utilizadas na maioria das redes neurais para reconhecimento de imagem (DATA SCIENCE ACADEMY, 2021). Trabalhos anteriores propõem o emprego de redes neurais convolucionais, aliadas à técnica de transferência de aprendizado, para a realização das tarefas de detecção e classificação materiais recicláveis (KULKARNI, RAMAN, 2019; HE, GU, SHI, 2020; DEWULF, 2017).

A dificuldade de se encontrar grandes bases de dados com imagens de materiais recicláveis em estado de pós-uso devidamente rotulados, representa um obstáculo no treinamento da CNN. Em Kulkarni, Raman, 2019 e Dewulf, 2017, os autores utilizaram técnicas de *data augmentation* e de transferência de aprendizado para minimizar o problema, obtendo acurácias na ordem de 80%-90% nos conjuntos de teste. No entanto, não puderam testar suas redes em um dataset que se aproxime mais à realidade da disposição dos materiais nas esteiras de triagem manual.

Recentemente, foi disponibilizada a base de imagens ZeroWaste (BASHKIROVA et al., 2021), a qual contém 1800 imagens segmentadas rotuladas, além de 6000 imagens não rotuladas, obtidas a partir de quadros de vídeos coletados da esteira de triagem de uma unidade real de processamento de materiais reciclados.

1.2 Justificativa e motivação

O processo de reciclagem é uma etapa fundamental da estruturação de um plano de gestão de resíduos sólidos sustentável. Estima-se que a produção global de resíduos sólidos urbanos aumentará em 70% até 2050, com a disseminação e otimização do processo de reciclagem, pode-se reintroduzir parte desses resíduos na cadeia produtiva. Essa reintrodução diminui a dependência da sociedade com relação às matérias-primas brutas e diminui o montante final de resíduos que precisam ser alocados em aterros sanitários (KAZA et al., 2018).

A separação dos diferentes materiais recicláveis é a chamada etapa de triagem, sendo essa o gargalo do processo (PARREIRA, OLIVEIRA, LIMA, 2009). A utilização de técnicas de visão computacional foi proposta por diversos autores (KULKARNI, RAMAN, 2019; HE, GU, SHI, 2020; DEWULF, 2017) para auxiliar a otimizar esta tarefa. Mais especificamente, o uso de CNN's, as quais são amplamente utilizadas em tarefa de detecção e classificação de imagens (DATA SCIENCE ACADEMY, 2021).

No entanto, não se encontrou um trabalho anterior que tenha testado essa abordagem com dados reais, advindos da esteira de triagem de uma planta de reciclagem. Este projeto visa sanar esta falta, buscando verificar a viabilidade de implantação de sistemas inteligentes nas linhas de produção, otimizando sua eficiência.

1.3 Questão de pesquisa e objetivos

Neste trabalho, espera-se avaliar o desempenho de redes convolucionais nas tarefas de detecção e classificação de imagens do dataset ZeroWaste (BASHKIROVA et al., 2021).

Questão: “Qual o desempenho obtido por diferentes implementações de uma mesma arquitetura de detecção e classificação de imagens estado-da-arte (KULKARNI, RAMAN, 2019; HE, GU, SHI, 2020; DEWULF, 2017) quando submetida a um banco de imagens que representa a realidade do processo de triagem em uma planta de reciclagem?”

Diante desta questão de pesquisa, são definidos os seguintes objetivos para o desenvolvimento deste trabalho:

- Mapear algoritmos de aprendizado a partir de redes convolucionais disponíveis na literatura;
- Treinar variações do algoritmo escolhido com o dataset ZeroWaste (BASHKIROVA et al., 2021), comparando os desempenhos obtidos para as diferentes composições.

A partir do modelo proposto, espera-se obter ao menos uma rede que apresente resultados satisfatórios.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Processo de reciclagem

O processo de reciclagem consiste na estratégia de reinserir materiais específicos, tais como papel, plástico, vidro, alumínio, os quais foram previamente descartados no processo produtivo através do reprocessamento adequado destes resíduos. Tal estratégia é fundamental para se estabelecer uma política de gerenciamento de resíduos sólidos urbanos bem-sucedida, a qual, por sua vez, é essencial à construção de uma sociedade mais sustentável, tanto social quanto ambientalmente (OZDEMIR et al., 2021).

O processo de reciclagem pode ser dividido em quatro grandes etapas (OZDEMIR et al., 2021):

- Coleta: esta etapa inclui diversas estratégias de coleta, como por exemplo, coletas seletivas em áreas residenciais, descartes de materiais em indústrias, centros de recepção de materiais coletados por agentes independentes, etc.;
- Processamento: esta etapa inclui a ida dos materiais aos centros de reciclagem, onde estes materiais serão separados, limpos e transformados em fonte de matéria-prima para processos produtivos posteriores;
- Produção: esta etapa consiste na realização de novos produtos a partir do material reciclado processado, tais como latas de alumínio, caixas de papelão, pregos, etc.;
- Venda: nesta etapa os produtos feitos a partir de materiais reciclados são vendidos, introduzindo circularidade ao modelo de consumo, tornando-o mais sustentável.

Figura 1: Principais etapas do processo de reciclagem e sua circularidade.

Fonte: autoria própria. Ícones por iconixar, Freepik e photo3idea_studio (www.flaticon.com).

Dentre as etapas descritas, a etapa de processamento, especificamente a etapa de separação dos materiais recicláveis, consiste na etapa mais crítica do processo de reciclagem (OZDEMIR et al., 2021). A separação é fundamental para se evitar contaminação dos materiais, assim como para aumentar a eficiência na geração de matéria-prima útil aos processos produtivos posteriores.

Existem três categorias de separação (triagem) de materiais (HADDAD et al., 2020):

- Manual: o material recebido é disposto sobre uma superfície (mesa ou chão) para então ser separado através da catação, sem uso direto de qualquer equipamento (pode haver presença de balanças, prensas e empilhadeiras na planta, mas não são utilizados na etapa de triagem);
- Semimecanizada: conta com a presença de esteiras transportadoras, as quais promovem um fluxo contínuo de materiais, que por sua vez serão separados através da catação de colaboradores dispostos ao longo da esteira. Pode contar com equipamentos adicionais, tais como extratores de sucata (separação magnética), no entanto, a triagem ainda é feita majoritariamente por pessoas;
- Mecanizada: na triagem mecanizada equipamentos, tais como peneiras rotativas, separadores balísticos, ópticos, eletrostáticos, indutivos e magnéticos, são responsáveis pela maior parte da separação. O papel dos colaboradores

passa a ser de inspeção final, na qual qualquer produto impróprio (erroneamente separado pelos equipamentos) é retirado da linha através de catação.

Figura 2: O processo de triagem manual é representado por **a** (chão) e por **b** (mesa). A triagem semiautomatizada pode ser observada em **c**, enquanto uma planta de triagem mecanizada é ilustrada em **d**.

Fonte: imagens retiradas de Britto (2018), Gomes (2018), Aversani (2020) e RNSP (2014).

Segundo Parreira et al. (2009) grande parte do material reciclado no Brasil é reintroduzido no ciclo produtivo graças às ações de associações e cooperativas de catadores. Existem ao menos 1,8 mil cooperativas atuantes no país atualmente, as quais reciclaram em média 510 toneladas de material em 2021 (ANCAT, 2021).

Britto (2018) levantou as condições de mecanização em associações e cooperativas capixabas, chegando à conclusão de que todas as entidades analisadas realizavam a triagem de forma semimecanizada (conta com presença de esteira para a triagem manual de resíduos) ou manual. Já um levantamento feito entre as cooperativas presentes no município de São Paulo e habilitadas pela Autoridade Municipal de Limpeza Urbana (AMLURB) apontou que, em 2014, nenhuma delas possuía um fluxo de trabalho mecanizado (ABLP, 2014).

Assumindo que as amostras de associações e cooperativas consideradas em ambos os estudos (BRITTO, 2018 e ABLP, 2014) são representativas da realidade nacional, pode-se concluir que no Brasil a maioria do material reciclado é processado em unidades de triagem manuais ou semiautomatizadas.

Visando comparar a eficiência produtiva de plantas de reciclagem com diferentes métodos de triagem, Haddad et al. (2020) levantou informações referentes à 6 cooperativas paulistanas, as quais foram sintetizadas na tabela 1.

Tabela 1: características de triagem de seis cooperativas de reciclagem paulistanas

	A	B	C	D	E	F
Triagem	Manual	Manual	Semimecanizada	Semimecanizada	Semimecanizada	Mecanizada
Material processado por cooperado $\left(\frac{ton}{mês}\right)$	4,00	2,78	4,46	3,66	4,69	22,0
Rejeito gerado (%)	30,0	5,0	10,0	8,5	12,0	50,0

Fonte 1: adaptado de Haddad et al. (2020).

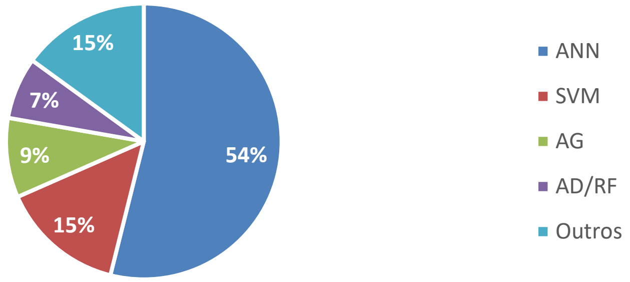
Pode-se perceber que a quantidade de materiais processados por cooperador em uma unidade mecanizada pode chegar a ser cinco vezes maior do que a média das demais cooperativas, no entanto a taxa de rejeito também cresce significativamente. Mesmo considerando o impacto da taxa da alta taxa rejeito da unidade mecanizada (50%), a taxa de material processado por cooperado em uma unidade mecanizada superior ao dobro da média para as demais plantas de separação.

Visando automatizar a triagem de materiais, mantendo a taxa de rejeito inferior à 50%, preferencialmente no limite de 10% (média para as cooperativas semimecanizadas em Haddad et al. (2020)), sugere-se a aplicação de técnicas de processamento de dados de imagens realizado através de modelos de aprendizagem profunda (*deep learning*).

2.2 Aprendizagem profunda aplicada à reciclagem

Guo et al., 2021 realizou a revisão de artigos publicados entre 2003 e 2020 os quais aplicassem técnicas de aprendizagem de máquina no contexto do tratamento de resíduos sólidos orgânicos e processos de reciclagem. Este levantamento apontou que em 54% dos artigos revisados redes neurais artificiais (*artificial neural networks* - ANN) foram o modelo escolhido para resolver o problema proposto, contra 15% SVM (*Single Vector Machine*), 9% algoritmos genéticos (AG), 7% de árvores de decisão e random forests (AD/RF).

Figura 3: Proporção de modelos de aprendizagem de máquinas utilizados em análises de tratamentos de resíduos e/ou processo de reciclagem.



Fonte: adaptado de (GUO et al., 2021).

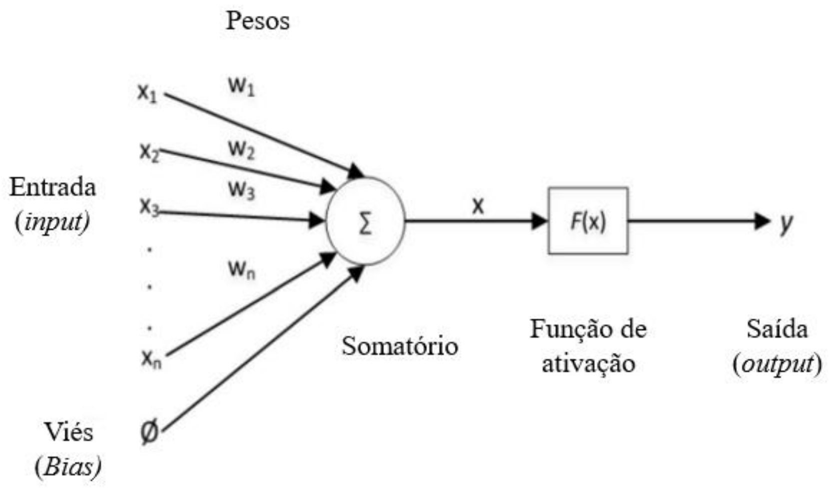
Por se tratar do modelo de aprendizado de máquina mais empregado na bibliografia específica pesquisada (GUO et al., 2021), as redes neurais artificiais são o alvo de estudo deste trabalho.

2.2.1 Redes neurais artificiais (ANNs)

Uma ANN consiste em uma unidade de processamento de informação cuja estruturação foi inspirada no modelo de funcionamento do cérebro humano (OZDEMIR et al., 2021). Algumas características típicas das ANNs são a sua não-linearidade, sua alta adaptabilidade e sua tolerância a falhas (GUO et al., 2021).

A unidade mínima de processamento de uma ANN recebe o nome de neurônio ou nó. O processamento de dados em um neurônio é realizado a partir da soma ponderada dos valores de entrada, a este valor adiciona-se também o viés adequado. O resultado desta soma é alimentado à função de ativação, a qual gerará uma determinada saída (figura 4).

Figura 4 – Representação da estrutura básica de um neurônio.



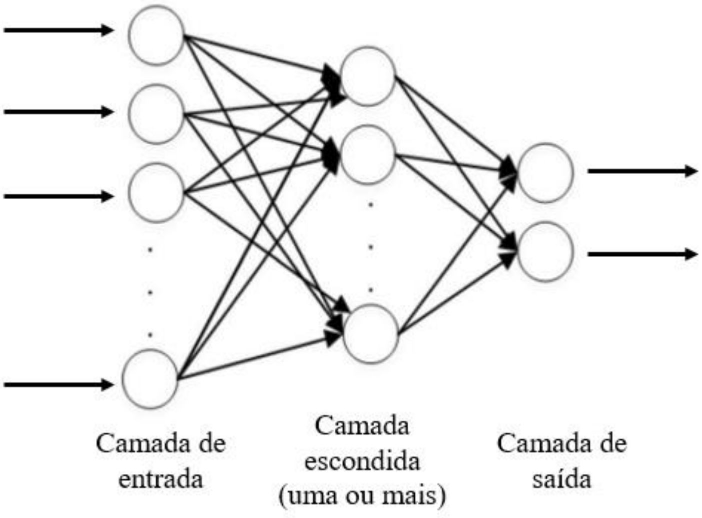
Fonte: adaptado de (OZDEMIR et al., 2021)

Geralmente em uma ANN múltiplos neurônios são utilizados, estes são organizados em camadas. Primeiramente temos a camada de entrada, a qual é seguida por uma ou mais camadas escondidas, e por último temos a camada de saída. Os neurônios de uma camada conectam-se com os neurônios da camada subsequente, de forma que as saídas dos neurônios de uma camada tornam-se as entradas dos neurônios da seguinte (figura 5). Como cada neurônio possui uma função de ativação, essas conexões sucessivas resultam em uma composição de funções, tal composição é o que caracteriza este modelo de aprendizagem como profundo (PONTI, COSTA, 2017).

Treinar um ANN implica em determinar valores ótimos para seus parâmetros (pesos e vieses). Para tal é necessário que se haja uma função que determine a qualidade da predição feita pela rede, tal função recebe o nome de função de custo. A função de custo indica quão longe uma determinada predição emitida pelo modelo está da classe real à qual a entrada analisada pertence. Uma das funções de custo mais utilizadas em classificação é chamada entropia cruzada (*cross-entropy*) (PONTI, COSTA, 2017).

Uma vez estipulada a função de custo utilizam-se algoritmos de otimização, os quais tem como objetivo encontrar o ponto de mínimo global da função de custo, para determinar os parâmetros ótimos do da rede neural. Gradiente descendente estocástico (*Stochastic Gradient Descent* – SGD) e Adam são alguns dos algoritmos de otimização utilizados na literatura (PONTI, COSTA, 2017).

Figura 5 – Representação da estrutura básica de uma ANN.



Fonte: adaptado de (OZDEMIR et al., 2021).

São exemplos de ANNs, perceptrons multicamadas (*multi layer perceptrons* – MLP), redes de função de base radial (*Radial Basis Function* – RBF), redes neurais convolucionais (*Convolutional Neural Networks* – CNN), redes neurais recorrentes (*recurrent neural network* – RNN).

Apesar das ANNs serem ferramentas úteis para lidar com grandes volumes de dados, alguns pontos de atenção devem ser levantados. A falta de explicabilidade dos resultados obtidos a partir das ANNs representa um problema. ANNs são por vezes consideradas “caixas pretas”, nas quais tem-se um input e um output claros, mas não há um entendimento completo das etapas intermediárias, tal característica pode fazer com que o output não seja percebido como confiável, principalmente em aplicações das ciências da natureza (GUO et al., 2021).

Outro ponto de atenção deve ser direcionado à alta adaptabilidade do modelo, apesar de termos mencionado esta característica como uma vantagem das ANNs, caso a modelagem não seja feita de forma adequada, a grande quantidade de parâmetros utilizados pode levar ao fenômeno de *overfitting*, fazendo com que o modelo final não desempenhe bem no ambiente real.

Neste trabalho, os dados que alimentarão o algoritmo de aprendizado de máquina consistem em imagens. O modelo de ANN mais consagrado para o processamento de imagens é a CNN (MAO et al., 2020; PONTI, COSTA, 2017), portanto este será o algoritmo alvo de nosso estudo.

2.2.2 Redes neurais convolucionais (CNNs)

Uma CNN é uma rede neural composta basicamente por camadas convolucionais. Tais camadas processam as entradas a partir de campos receptivos locais. A convolução permite processar imagens de entrada levando em conta sua estrutura bidimensional, tal característica faz com que a principal aplicação das CNNs seja o processamento de informações visuais (PONTI, COSTA, 2017).

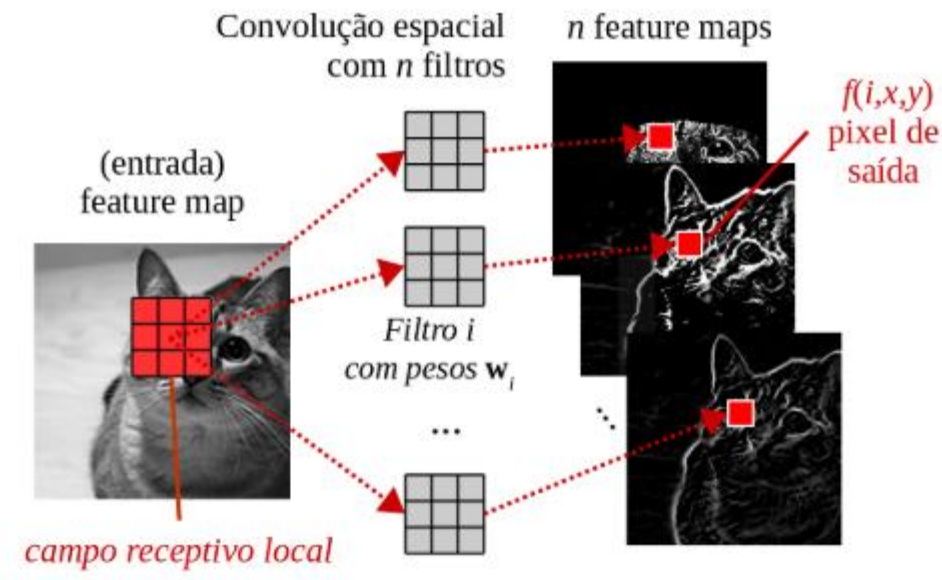
Cada neurônio de uma camada convolucional consiste em um filtro aplicado a uma imagem de entrada. Este filtro é composto por um tensor (matriz multidimensional) de pesos, o qual é responsável pela transformação da imagem de entrada por meio de uma combinação linear dos pixels vizinhos (PONTI, COSTA, 2017).

A dimensão do tensor filtro é dada por $k \times k \times d$, onde k é parâmetro a ser definido pelo designer do modelo, enquanto d é dado a partir do número de canais de entrada, por exemplo, uma imagem RGB possui 3 canais, portanto os filtros aplicados a essa imagem deverão ter $d = 3$. Os tamanhos de filtro mais utilizados são $5 \times 5 \times d;$ $3 \times 3 \times d$ e $1 \times 1 \times d$ (PONTI, COSTA, 2017).

Esta estrutura de processamento através de filtros faz com que haja uma redução significativa da quantidade de pesos do modelo, por exemplo, para uma imagem RGB de tamanho $224 \times 224 \times 3$, 150528 pesos seriam necessários em um neurônio completamente conectado (*fully connected* – FC), enquanto para um filtro convolucional com $k = 5$, teríamos $5 \times 5 \times 3$, 75 pesos em um mesmo neurônio (PONTI, COSTA, 2017).

Uma vez determinadas as dimensões do filtro, seus pesos e o viés do neurônio, este tensor é aplicado a uma região específica da imagem em processamento, esta região é chamada de campo receptivo local, cujo valor de saída (pixel) é dado pela combinação dos campos de entrada nesse campo receptivo (PONTI, COSTA, 2017). A saída resultante deste processo consiste em uma nova matriz de pixels, a qual recebe o nome de *feature maps* (figura 6).

Figura 6 – Representação esquemática de uma camada convolucional com n filtros processando uma imagem em preto e branco (como a imagem possui um único canal, temos $d = 1$).



Fonte: PONTI, COSTA, 2017.

Para $k = 3$, o campo receptivo é composto por 9 pixels vizinhos. No exemplo apresentado na figura 7, temos como primeiro campo receptivo a ser processado a matriz $\begin{bmatrix} 2 & 2 & 2 \\ 1 & 0 & 1 \\ 1 & 1 & 3 \end{bmatrix}$ (em destaque na matriz de entrada). Para se determinar o pixel de saída deste campo receptivo, um neurônio realizará o produto da matriz do campo receptivo pela matriz do filtro e adicionará o viés ao resultado obtido (equação 1).

Figura 7 – Exemplo de como é determinado o pixel de saída de um campo receptivo.

Entrada (7 x 7 x 1)

	0	1	2	3	4	5	6
0	2	2	2	2	3	3	3
1	1	0	1	1	1	1	0
2	1	1	3	3	0	0	0
3	1	1	3	2	0	0	3
4	1	1	3	2	0	0	3
5	1	3	3	2	0	0	3
6	3	3	3	2	0	0	3

Filtro
(3 x 3 x 1)

-1	0.5	1
-1	0	0
0	0	0.5

Viés

1.5

Saída (5 x 5 x 1)

	0	1	2	3	4
0	3				
1					
2					
3					
4					

Fonte: adaptado de Ponti, 2021.

Equação 1 – Primeiro passo do cálculo realizado pelo neurônio de convolução

$$Pixel\ de\ saída_{0,0} = \begin{bmatrix} 2 & 2 & 2 \\ 1 & 0 & 1 \\ 1 & 1 & 3 \end{bmatrix} \times \begin{bmatrix} -1 & 0.5 & 1 \\ -1 & 0 & 0 \\ 0 & 0 & 0.5 \end{bmatrix} + 1.5 = 3$$

Adicionalmente aos pontos de atenção levantados na seção anterior para as ANNs, quando da utilização de CNN’s deve-se observar que como o único input para extração de *features* são imagens, a qualidade do resultado é extremamente sensível à qualidade das imagens que alimentam o modelo (GUO et al., 2021).

2.2.3 Classificação multirrótulos

As imagens contidas no dataset de interesse possuem uma diferença fundamental com relação à TrashNet: presença de múltiplos objetos sujeitos à classificação em cada uma das imagens.

Surge a partir dessa característica a necessidade de uma camada adicional de processamento que seja capaz de localizar os objetos classificáveis dentro de uma imagem, para então dar prosseguimento à tarefa de classificação.

Figura 8 – Exemplos de imagens que contém múltiplos rótulos obtidos a partir de (a) detecção de objetos e (b) segmentação de objetos

Fonte: adaptado de BANDYOPADHYAY, 2022

Para a classificação de imagens com múltiplos rótulos agregados a localização de objetos, existem duas principais vertentes de localização (BANDYOPADHYAY, 2022): detecção de objetos e a segmentação de imagens. No caso da detecção de objetos, a localização dos objetos é apontada através de *bounding boxes*, em geral com formato retangular (figura 8a). Já no caso da segmentação de imagens, cada pixel da imagem será classificado como pertencente a um objeto ou classe de objetos, gerando máscaras que delimitam os contornos dos objetos (figura 8b).

Para o caso em estudo, o objetivo final da correta classificação dos materiais recicláveis é a possibilidade de automatizar o processo a partir, por exemplo, de braços robóticos, para tal, seria imprescindível uma visão computacional clara do contorno dos objetos. A densidade dos objetos na esteira de separação é considerada alta, o que também sugere que os esforços de localização de objetos devem ser direcionados no sentido da correta segmentação das imagens obtidas pelos sensores. Portanto, opta-se pelo emprego da segmentação de imagens para a classificação dos objetos contidos no dataset ZeroWaste.

2.2.3.1 Segmentação de imagens

Tarefas de segmentação de imagens podem ser classificadas em três categorias (SALMI, 2021):

- Segmentação semântica
- Segmentação de instâncias
- Segmentação panóptica

Para entender a diferenciação entre elas, primeiro é necessário definir os conceitos de "*stuff*" e "*things"* em uma imagem. *Things* contempla todas as categorias de objetos contáveis contidos em uma imagem, tais objetos podem ser contados na imagem ao se atribuir diferentes Ids para cada uma dessas instâncias. *Stuff* representa todas as classes de incontáveis, tais como céu, estrada, mar, etc (V7 Labs, 2022).

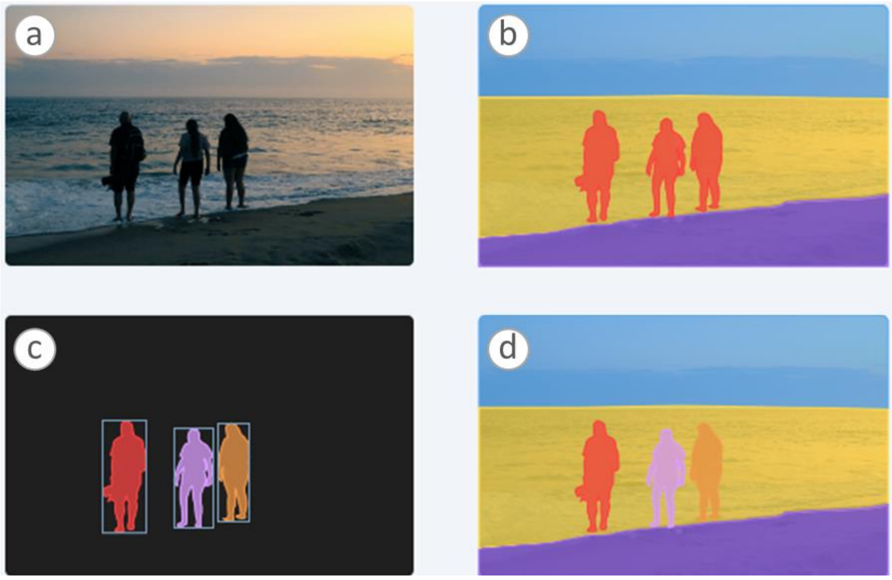
A segmentação semântica realiza a classificação de todos os pixels de uma imagem, sendo que para cada um deles é atribuída uma classe dentre as predefinidas pelo usuário.

Já a segmentação de instâncias geralmente gera *bounding boxes* que delimitam cada objeto contável (*thing*) presente na imagem, juntamente com sua respectiva máscara e classificação, nesse caso, múltiplos objetos pertencentes a uma mesma classe são tratados como instâncias distintas.

Por fim, a segmentação panóptica utiliza um algoritmo capaz de diferenciar diferentes objetos de uma mesma classe (segmentação de instâncias), sendo também capaz de classificar os objetos não contáveis.

A diferenciação entre as saídas de uma imagem submetida a cada uma das três vertentes de segmentação pode ser observada na figura 9.

Figura 9 – Diferentes métodos de segmentação de imagens: (a) imagem original; (b) segmentação semântica; (c) segmentação de instâncias; (d) segmentação panóptica



Fonte: adaptado de V7 Labs, 2022.

O ZeroWaste dataset possui anotações somente para as instâncias de interesse (plástico, plástico rígido, papelão e metal), não havendo máscaras para papéis (pois não devem ser removidos da esteira de separação), assim como para o fundo da imagem (esteira de separação). Portanto, o método de segmentação mais adequado consiste na segmentação e instâncias, assim como sugerido no artigo de publicação da base de imagens.

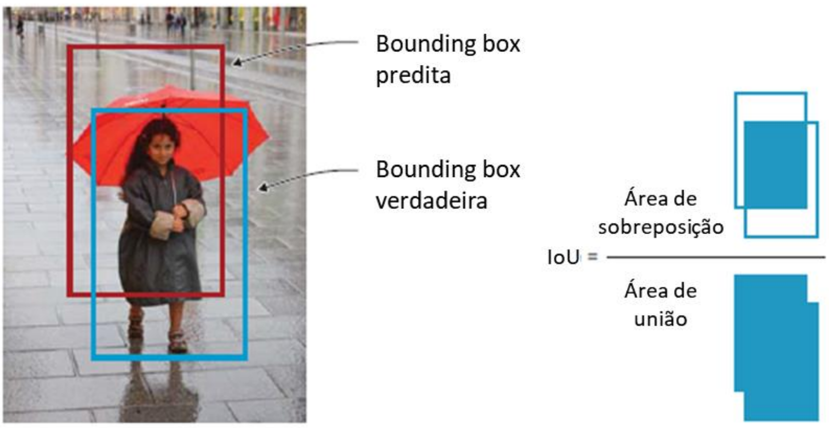
Usualmente frameworks de detecção de objetos contém quatro componentes (ELGENDY, 2020):

- *Region proposal*: modelo utilizado para gerar regiões de interesse dentro de uma imagem (ROI – *Regions of Interest*) que serão efetivamente processadas para geração das máscaras e classificações de cada objeto. A saída deste modelo consiste em uma grande quantidade de *bounding boxes* (coordenadas que definem um contorno, geralmente retangular), cada qual com uma respectiva pontuação de objetividade (*objectness score*). As *bounding boxes* com maior pontuação são então repassadas para as camadas de processamento seguintes.
- Extratores de características e preditores: para cada *bounding box* características são extraídas, a partir delas os preditores constatam a presença ou não de um objeto no contorno, e caso a presença seja confirmada, é realizada a classificação do objeto
- *Non-maximum suppression (NMS)*: usualmente múltiplas *bounding boxes (bbox)* são fornecidas para uma mesma imagem, sendo frequente a sobreposição das mesmas, portanto o

objetivo da técnica de NMS é realizar a combinação de *bbox* sobrepostas, resultando em uma única bbox para cada uma das instâncias de interesse

- Métricas de avaliação de desempenho: as métricas utilizadas para medir o desempenho da tarefa de detecção de objetos são:
    - Curva de precisão e recall: assim como para classificação de imagens, a curva de precisão e recall consiste na plotagem de uma curva que possui a precisão (razão entre os verdadeiros positivos detectados pelo algoritmo e a soma dos verdadeiros positivos com os falsos positivos) no eixo y e o recall (razão entre os verdadeiros positivos detectados pelo algoritmo e a soma dos verdadeiros positivos com os falsos negativos) no eixo x. Um bom detector consiste naquele que mantém índices de precisão altos conforme o recall aumenta.
    - Quadros processados por segundo (*FPS – frames per second*): determina a velocidade de processamento de um determinado framework.
    - Precisão média (*mAP – mean average precision*): uma das métricas mais utilizadas na tarefa de detecção de objetos, consiste em uma porcentagem, calculada a partir da média entre as áreas sob as curvas de precisão x recall para todas as classes existentes, sendo que quanto maior a mAP, melhor o resultado obtido pelo algoritmo.
    - Interseçcão sobre união (*IoU – intersection over union*): determina a sobreposição entre a bbox predita e a bbox verdadeira. Tal medida é utiliza para determinar se a detecção prevista é um positivo verdadeiro ou falso, sendo calculada a partir da razão entre a área de sobreposição entre as bboxes e a área de união das mesmas (figura 10). Quanto maior a razão, maior a qualidade da bbox predita, no entanto, usualmente utiliza-se um valor mínimo de IoU=0,5 para que a predição seja considerada como um verdadeiro positivo.

Figura 10: Ilustração para definição da métrica de intersecção sobre união (IoU)



Fonte: adaptado de Elgendy, 2020

Três arquiteturas populares para a tarefa de detecção de objetos são as redes neurais convolucionais baseadas em regiões (*region-based convolutional neural networks – R-CNN*), *single-shot detectors (SSD)* e YOLO (*you only look once*).

2.2.3.2 Mask R-CNN

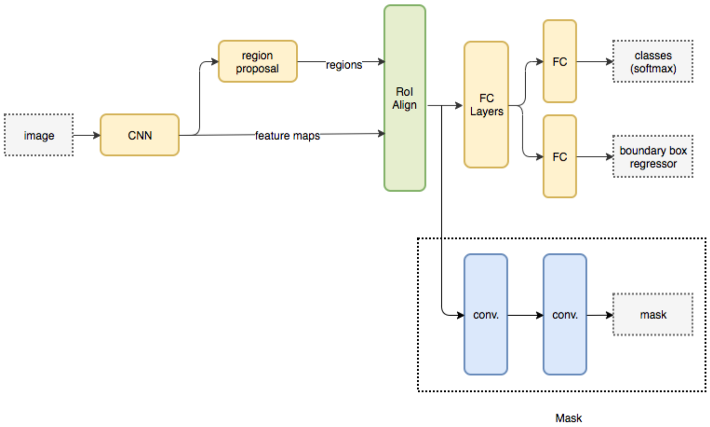
A arquitetura Mask R-CNN é uma das atualizações mais recentes da família R-CNN, sendo construída sobre o modelo Faster R-CNN com um ramo adicional para a segmentação dos objetos de interesse contidos em uma imagem.

Resumidamente, a estrutura geral do MASK R-CNN pode ser observada na figura 11. A primeira etapa de processamento consiste na extração de mapa de características a partir de uma rede neural convolucional (*backbone*), tais mapas são então alimentadas a um *region proposal*, que gera como saída as coordenadas das bboxes assim como a pontuação de objetividade para cada uma delas.

Ambas as informações, juntamente com os mapas de características são então encaminhados para a uma cada de de pooling das regiões de interesse (RoI), cuja saída consiste em RoI de tamanho fixo. Sequencialmente tais RoI são processadas por camadas completamente conectadas (FC) até chegarem às duas últimas camadas conectadas, uma dedicada a classificação do objeto, e outra dedicada a identificação das coordenadas da bbox associada a ele.

O ramo adicional responsável pela geração das máscaras de segmentação das instâncias de interesse recebe o output da camada de pooling das RoI, o qual é processado por camadas convolucionais, gerando como resultado a máscara desejada.

Figura 11 – Diagrama simplificado representando a arquitetura do modelo Mask R-CNN



Fonte: Hui, 2018.

2.2.3.3 SSD

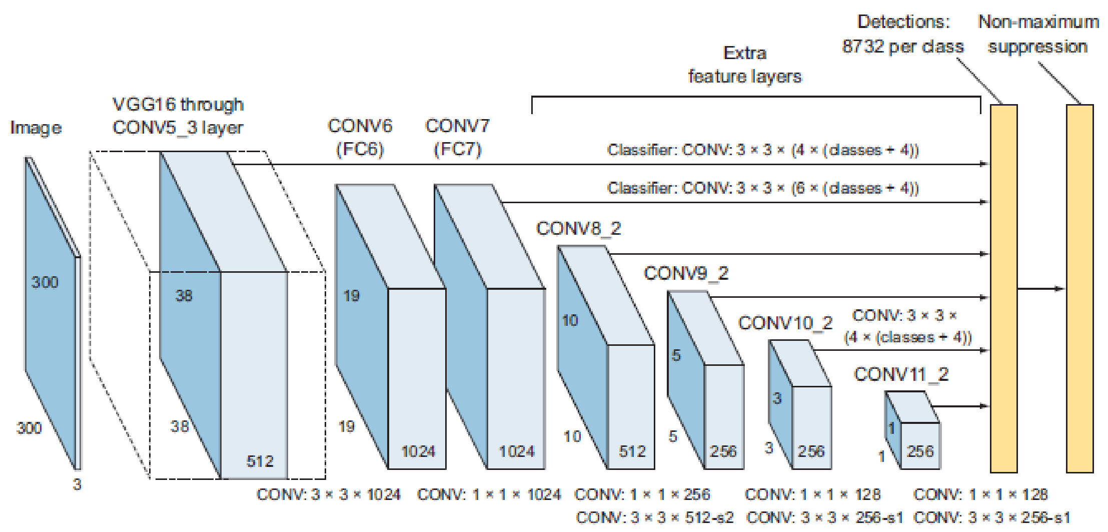
Ao contrário da família R-CNN, SSD e YOLO são detectores com um uma única fase de processamento, isto é, ao contrário das R-CNNs nas quais existe uma rede que avalia a localização do objeto e outra que realiza sua classificação, ambas as tarefas são realizadas por camadas convolucionais.

Desta forma, em detectores de fase única a pontuação de objetividade é determinada a partir de regressão logística para cada bbox, caso a pontuação obtida seja maior que um limiar pré-determinado, então o modelo realiza uma classificação, caso contrário, a etapa de predição é dispensada.

Tal característica permite que estes detectores obtenham FPS superiores àqueles obtidos por R-CNN's, em detrimento de uma menor mAP.

A arquitetura básica de uma SSD pode ser visualizada na figura 12. Consiste basicamente em uma rede neural convolucional utilizada como base (*backbone*) a partir da qual os mapas de características são extraídos. Posteriormente, camadas com múltiplas escalas decrescentes são adicionadas, por fim, uma camada de NMS, a qual recebe bboxes de camadas convolucionais de diferentes escalas, é responsável pela eliminação de bboxes sobrepostas e pela classificação dos objetos.

Figura 12: SSD com VGG16 como backbone. Neste exemplo, a penúltima camada recebe 8732 bboxes, os quais resultam da conexão da camada conv4_3 (38 x 38 x 4 bbox), conv7 (19 x 19 x 6 bboxes), conv8_2 (10 x 10 x 6 bboxes), conv9_2 (5 x 5 x 6 bboxes), conv10_2 (3 x 3 x 4 bboxes) e conv11_2 (1 x 1 x 4 bboxes)



Fonte: Elgendy, 2020

A saída esperada da camada de detecção consiste em vetores de dimensão 5 (4 coordenadas da bbox + 1 pontuação de objetividade) + número de classes do problema, tal vetor é repassado para a NMS, a qual é responsável por ranquear as predições e manter apenas a quantidade desejada (geralmente, no máxima as 200 melhores predições (ELGENDY, 2020)).

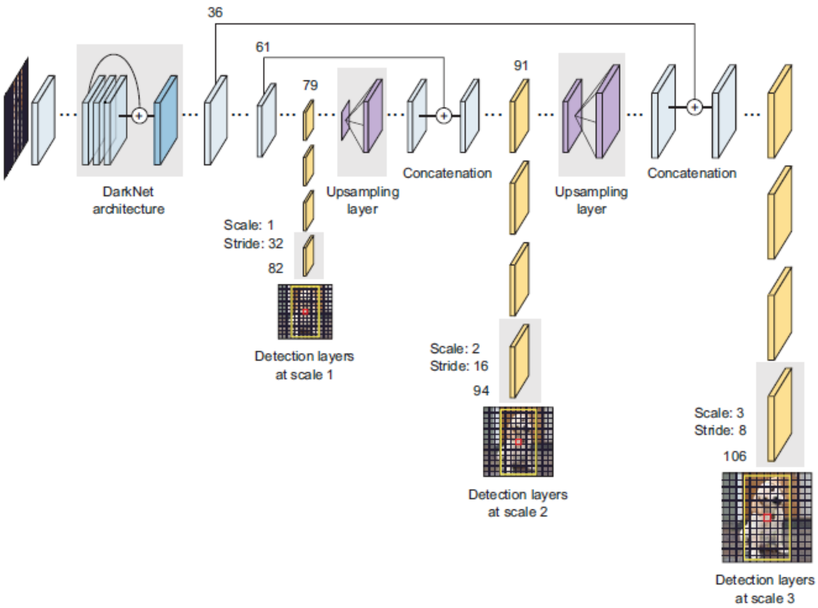
2.2.3.4 YOLO

Assim como a SSD, a YOLO consiste em um detector de fase única. Uma de suas versões mais recentes é a YOLOv3, cuja arquitetura pode ser observada na figura 13.

A YOLO prevê três momentos distintos de detecção, visando realizá-la em três diferentes escalas com passos 32 (detecção de objetos grandes), 16 (detecção de objetos médios) e 8 (detecção de objetos pequenos).

Inicialmente uma imagem é processada pela DarkNet-53, a qual conta com 53 camadas, em seguida a imagem sofre *downsampling* até a camada 79, a partir da qual a rede se ramifica, e prossegue com a redução, até a primeira detecção ser feita na camada 82 (passo 32). Em seguida o mapa de características da camada 79 é *upsampled* por 2 e concatenado com o mapa advindo da camada 61, então uma segunda predição é realizada na camada 94 (passo 16). Processo semelhante ocorre para as camadas subsequentes, de forma que a última predição é realizada na camada 106 (passo 8).

Figura 13: arquitetura completa da YOLOv3



Fonte: Eldengy, 2020

2.3 Estado da Arte

2.3.1 Classificação de imagens

A inserção de técnicas de aprendizagem profunda no universo de gerenciamento de resíduos sólidos urbanos é um fenômeno recente, nota-se que a maioria dos artigos sobre o tema datam a partir de 2018.

Como a otimização do desempenho dos modelos de aprendizagem profunda depende da exposição do modelo a uma grande quantidade de dados, a escassez de bases de imagens especificamente voltadas aos RSUs representa um obstáculo ao emprego das CNNs.

Yang e Thung (2016) publicaram uma base de imagens conhecida como TrashNet. A TrashNet conta atualmente com um total de 2527 imagens rotuladas em seis diferentes categorias: vidro (501), papel (594), papelão (403), plástico (482), metal (410) e não reciclável (137). As imagens consistem em fotos individuais de cada peça de material reciclado sobre um fundo branco, expostos à iluminação natural.

A TrashNet deu o primeiro passo para sanar o déficit de bancos de dados de imagens de domínio específico no campo da reciclagem, a maioria dos artigos aqui referenciados utilizaram imagens contidas nessa base como entrada nos seus modelos de classificação.

Apesar de sua importância, a quantidade de imagens contidas na TrashNet é considerada pequena para o treinamento de uma CNN. Para contornar este problema, técnicas de *data augmentation*, transferência de aprendizado (*fine-tuning* e/ou *feature extraction*), e comitês de classificação (*ensemble*) são comumente vistas em publicações da área (tabela 2).

Sidhart (2020) construiu uma CNN cuja estrutura é composta por 3 camadas convolucionais com 32 filtros cada, cada uma delas seguida de MaxPooling2D (2×2). As *feature maps* resultantes são então achatadas e alimentada à duas camadas completamente conectadas com 128 neurônios, cujo resultado alimenta a camada de saída com 4 neurônios com função de ativação Softmax. O banco de dados utilizados é composto pelas imagens contidas em quatro das seis categorias da TrashNet (YANG, THUNG, 2016), papel, metal, plástico e papelão, totalizando 2077 imagens. A acurácia obtida para o conjunto de teste após 100 épocas foi de 76,19%. O resultado abaixo da média, quando comparado as demais referências (tabela 2), deve-se à maior simplicidade do modelo adotado.

Aral et al. (2018) optou por comparar o desempenho de arquiteturas mais robustas e já bem estabelecidas, evidenciando o impacto da transferência de aprendizado nas acurácias obtidas para a tarefa de classificação das imagens na TrashNet (YANG, THUNG, 2016). Técnicas simples de *data augmentation*, tais como espelhamento vertical e horizontal e rotações de 15º ou 20º, foram utilizadas para enriquecer a base de dados. Os melhores resultados para acurácia foram obtidos a partir das arquiteturas DenseNet121 (95%), DenseNet169 (95%) e Inception-V4 (94%).

Özkaya (2018) comparou o desempenho da tarefa de classificação ao substituir a camada de saída com função de ativação SoftMax por um classificador SVM. Todas as arquiteturas testadas (AlexNet, GoogleNet, ResNet, VGG-16 e SqueezeNet) foram pré-treinadas. O resultado indicou que a acurácia dos modelos conectados com SVM na tarefa de classificação da TrashNet (YANG, THUNG, 2016) foi superior aos modelos com Softmax. Destaca-se a GoogleNet+SVM, a qual atingiu o patamar de 97.86% de acurácia após 200 épocas.

Similarmente ao que foi feito por Özkaya (2018), Ramsurrun (2021) também optou por comparar diferentes classificadores na camada de saída: SVM, Softmax e Sigmoid. Processos de *data augmentation* foram empregados, no entanto, as arquiteturas não foram pré-treinadas. Inception-V4, Inception-V3 e ResNet101V2 são exemplos de arquiteturas que desempenharam melhor com o emprego da função Sigmoid; enquanto VGG-16 e MobileNet desempenharam melhor com o SVM; finalmente, VGG-19, Xception desempenharam melhor com o Softmax. Todas as arquiteturas foram submetidas à 50 épocas de treinamento, posteriormente as cinco melhores foram submetidas à 100 épocas de treinamento, como resultado a melhor acurácia na base de imagens TrashNet (YANG, THUNG, 2016) foi obtida a partir da estrutura VGG-19+Softmax (87.9%).

Huang et al. (2020) propõe a construção de um comitê de classificação composto por três arquiteturas como método para maximizar a acurácia da classificação das imagens contidas no banco de imagens TrashNet (YANG, THUNG, 2016). Quando combinadas as arquiteturas pré-treinadas VGG19 (89.7% de acurácia), DenseNet169 (88.6% de acurácia) e NASNetLarge (89.2% de acurácia), forma-se um comitê cuja acurácia de classificação atinge 96.5%.

Mao et al. (2020) propõe a utilização de algoritmos genéticos para otimizar os hiper parâmetros da camada completamente conectada (FC) da arquitetura DenseNet121. Tal otimização juntamente com técnicas de data augmentation aplicadas à TrashNet (YANG, THUNG, 2016) viabilizaram uma acurácia de classificação de 99.6%.

Vo et al. (2019) utilizou a arquitetura ResNext como base do seu modelo intitulado *Deep Neural Networks for Trash Classification* (DNN-TC). As modificações feitas consistem em adicionar duas camadas totalmente conectadas após a etapa de *Global Average Pooling* existente na ResNext-101, cujas dimensões de saída são 1024 e N respectivamente, sendo n o número de classes. Para o caso da TrashNet (YANG, THUNG, 2016), N=6. A função de ativação log softmax é empregada na camada de saída, gerando uma acurácia de 94% após 100 épocas de treinamento.

Por fim, Bircanoglu et al. (2018) optou por adaptar a arquitetura DenseNet121 de forma a diminuir sua complexidade, permitindo que o modelo seja treinado em hardwares mais simples. A RecycleNet possui 3 milhões de parâmetros, uma redução de mais de 50% com relação aos 7 milhões de parâmetros de seu modelo base. Apesar de atingir uma menor acurácia (81% na TrashNet (YANG, THUNG, 2016)), a RecycleNet diminui as restrições de hardware necessárias para treinar uma rede com 121 camadas.

Tabela 2 – Resumo dos resultados disponíveis na bibliografia para diferentes arquiteturas de CNNs

Referência	Nº de classes	Nº de imagens	Data augmentation	Transferência de aprendizado	Acurácia	Observações
Sidhart et al., 2020	4	2.1k	-	-	76.19%	100 épocas
Aral et al, 2018	6	2.5k	+	+	95%	DenseNet121 100 épocas
					95%	DenseNet169 120 épocas
					94%	Inception-V4 120 épocas
Özkaya e Seyfi, 2018	6	2.5k	-	+	97.86%	GoogleNet + SVM 200 épocas
			-	+	88.10%	GoogleNet + Softmax 200 épocas
			-	+	97.46%	VGG-16+SVM 200 épocas
			-	+	90%	VGG-16+Softmax 200 épocas
Ramsurrun et al., 2021	6	2.5k	+	-	87.9%	VGG19 + Softmax 50 épocas
Huang et al., 2020	6	7.5k	-	+	96.5%	VGG19 + DenseNet169 + NASNetLarge
Mao et al., 2020	6	2.5k	+	+	99.6%	DenseNet121 + Algoritmo genético 40 épocas
Vo et al., 2019	6	2.5k	-	+	94%	ResNext modificada 100 épocas
Bircanoglu et al., 2018	6	2.5k	+	+	81%	RecycleNet 200 épocas

Além da TrashNet (YANG, THUNG, 2016), outra base de imagens de materiais recicláveis disponível para download é a Recycling Dataset (SINGH, LUO, LI, 2021). A Recycling Dataset é composta por 11500 imagens divididas em cinco classes: caixas, garrafas de vidro, latas de bebidas, latas de bebidas amassadas e garrafas plástica, cada classe conta com 2300 imagens. Assim como a TrashNet, as imagens da Recycling Dataset possuem fundo claro, apresentando apenas um objeto por imagem.

2.3.2 Detecção de objetos

A escassez de bases de dados anotadas com múltiplos objetos recicláveis pós-descarte (com deformação e sobreposição significativa entre as instâncias) é um obstáculo para o desenvolvimento e comparação de arquiteturas de aprendizagem profunda que visem detectar e classificar estes materiais.

Visando contornar este problema Kulkarni e Raman (2019) realizaram um trabalho interessante de sobreposição dos objetos contidos em imagens da TrashNet, de forma a se aproximar da realidade de um ambiente de triagem real. Quatro objetos são recortados e colados em uma mesma imagem utilizando uma rede adversária generativa (*Generative Adversarial Networks* – GAN*)*, as imagens com múltiplos objetos são então alimentadas a uma rede do tipo Faster R-CNN, obtendo *F1-Score* de 0.98 para a classe "papelão" e 0.78 para a classe "não reciclável".

No entanto, Seredkin et al., 2019 notou que ao se montar uma base de dados de imagem sintética a partir de recortes retangulares de imagens que continham apenas um material reciclável, apesar do mesmo fundo uniforme ser utilizado para todas as imagens, o detector resultante apresentou bbox coincidentes com os exatos retornos dos recortes, indicando que uma possível diferença entre a iluminação das colagens estava sendo utilizada pela rede para amparar duas predições.

Uma série de outros bancos de dados de imagens de resíduos foram levantados durante o projeto *Detect Waste Project* (sem fins lucrativos), tais como Open Liter Map, TrashCan 1.0, Extended TACO, cujo framework de processamento proposto consiste em dividir o problema de detecção e classificação em duas partes, para cada uma delas uma rede neural dedicada é utilizada. Para o problema de detecção foram analisadas três redes: EfficientDet, DETR e Mask R-CNN, sendo que a melhor mAP foi aquela obtida pela EfficientDet (65,5%). Já para a etapa de classificação a rede EfficientNet-B2 foi utilizada por proporcionar melhores resultados quando comparada com a ResNet-50 e EfficienteNet-B4 (MAJCHROWSKA et al., 2021).

Em Majchrowska et al., 2021 ainda é descrito que a maioria das aplicações de aprendizagem profunda em tarefas de reconhecimento de imagens de RSU's utiliza redes da família R-CNN, SSDs e YOLO, com mAP variando entre 15,9% para a base de dados TACO com arquitetura Mask R-CNN e 81% para Trash-ICRA19 com a Faster R-CNN (tabela 3). Embora, na maioria das pesquisas, a quantidade de classes seja reduzida (geralmente apenas uma categoria do tipo “resíduo urbano”).

Tabela 3: resumo dos resultados disponíveis na bibliografia para diferentes arquiteturas de detectores de objetos

Referência	Base de dados	Arquitetura	Desempenho (%)
Awe, Mengistu, and Sreedhar 2017	Mindy Yang and Gary Thung’s dataset	Faster R-CNN	mAP = 68.3
Liu et al. 2018	VOC2007 dataset	YOLOv2	Acc = 89.71
Fulton et al. 2019	Trash-ICRA19	Faster R-CNN	mAP = 81
Fulton et al. 2019	Trash-ICRA19	YOLOv2	mAP = 47.9
Fulton et al. 2019	Trash-ICRA19	SSD	mAP = 67.4
Hong, Fulton, and Sattar 2020	TrashCan 1.0	Faster R-CNN	AP = 34.5
Hong, Fulton, and Sattar 2020	TrashCan 1.0	Mask R-CNN (segmentation)	AP = 30
Carolis, Ladogana, and Macchiarulo 2020	TrashNet	YOLOv3	mAP50 = 59.57
Proença and Simões 2020	TACO	Mask R-CNN (instance segmentation)	mAP50 = 15.9
Kraft et al. 2021	UAVVaste	YOLOv4	mAP = 47.6
Kraft et al. 2021	UAVVaste	EfficientDet-D3	mAP = 44.5
Liang and Gu 2021	WasteRL	ATSS	mAP = 67.5

Fonte: adaptado de Majchrowska, 2021.

3 METODOLOGIA

Embora, como levantado na seção anterior, pesquisas visando a aplicação de técnicas de aprendizagem profunda tem se tornado mais frequentes nos últimos anos, aplicações em ambientes produtivos ainda não são amplamente exploradas. A falta de uma base de dados de referência pública, devidamente anotada é um dos obstáculos ao desenvolvimento de arquiteturas que atendam a este domínio específico.

Visando sanar este déficit de bases de dados anotadas que contemplem a complexidade inerente ao processo de separação de RSUs, em 2021 foi disponibilizada a ZeroWaste Dataset (BASHKIROVA et al., 2021), a qual apresenta 1874 imagens completamente segmentadas e rotuladas de materiais reciclados em ambiente produtivo, mais especificamente na esteira de separação de uma unidade de triagem de RSU recicláveis. A base também contém outras 6212 imagens não rotuladas, as quais podem ser utilizadas, segundo Bashkirova et al. (2021), para treinar algoritmos semi-supervisionados.

Na esteira de triagem representada pela base de imagens, a separação dos materiais é feita de acordo com 5 classes de materiais: papel, papelão, plástico, plástico rígido e metal. Após a separação, apenas papeis devem permanecer na esteira, por isso, apenas foram anotados objetos pertencentes às quatro demais classes.

Os autores realizaram experimentos para detecção de objetos utilizando as arquiteturas Mask R-CNN (pré-treinada a partir da MS COCO), RetinaNet e TridentNet, obtendo mAP de 34.9, 33.5 e 36.3 respectivamente (treinamentos supervisionados). O objetivo deste trabalho é investigar o impacto da alteração do backbone sobre o desempenho de uma arquitetura estado-da-arte especializada em segmentação (predição de máscaras para instâncias de objetos que não são classificados como papel), Mask R-CNN sobre o dataset ZeroWaste.

3.1 ZeroWaste Dataset

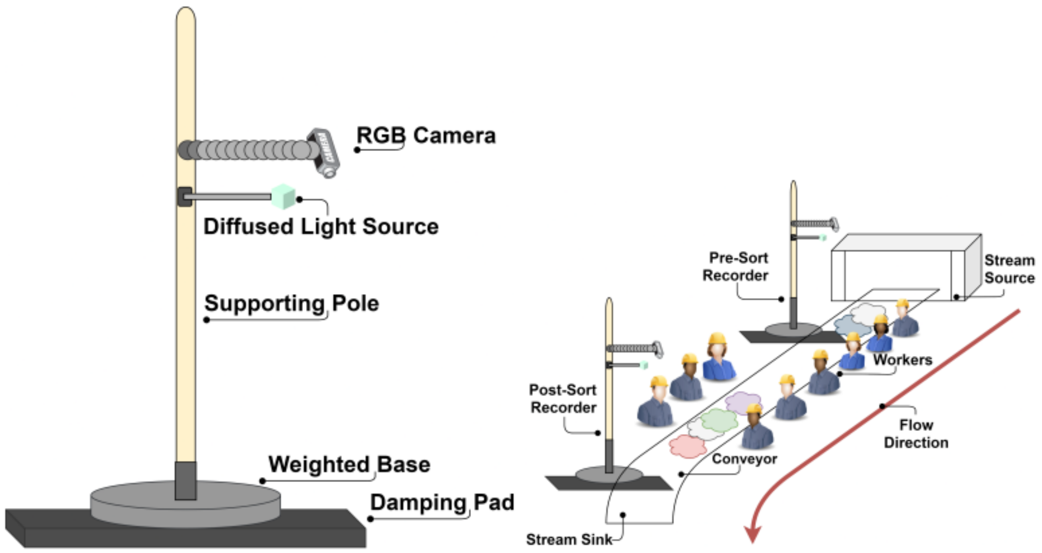
A ZeroWaste Dataset conta com quatro conjuntos de dados:

- ZeroWaste-f: banco de imagens anotadas, adotando-se o modelo COCO de anotação de bounding boxes e máscaras para cada instância de interesse presente
- ZeroWaste-w: coleção de dados no formato “antes x depois” da remoção dos objetos alvo de classificação, cujo objeto é fornecer dados para o treinamento de redes fracamente supervisionadas com saídas binárias

- ZeroWaste-s: banco de imagens não anotados, direcionado à métodos de aprendizagem semi-supervisionados
- ZeroWasteAug: base resultante da implementação de *data augmentation*, visando combater o desbalanceamento entre classes

As imagens foram coletadas utilizando o esquema apontado na figura 14.

Figura 14: configuração do aparato utilizado para obtenção da filmagem da esteira de separação. A esquerda: maiores detalhes do aparato de filmagem. A direita: disposição dos aparatos ao longo da esteira de separação



Fonte: Bashkirova, 2021

A planta de separação na qual a filmagem foi realizada é especializada em reciclagem de papel, de forma que qualquer outro material é considerado um contaminante (metal, plástico, etc.). A filmagem obteve imagens no início da esteira de separação e ao final dela (após a separação dos contaminantes). As câmeras (GoPro Hero 7) foram fixadas em bases desenhadas de forma a reduzir transmissão de vibração, visando a captura de quadros mais nítidos. Iluminação auxiliar foi providenciada graças ao uso de lâmpadas portáteis (LitraTorch 2.0) associadas a difusores de luz. Câmeras foram instaladas 1 metro acima da esteira de separação, e o ponto de iluminação a cerca de 0.8 m acima da mesma referência.

Foram obtidos 12 vídeos sequenciais com duração total de 95 minutos e 14 segundos, com 120 quadros por segundo e definição 1920 x 1080 pixels.

Os vídeos foram processados de acordo com as etapas abaixo indicadas, o resultado pode ser observado na figura 15:

1. Rotação e corte: quadros foram rotacionados de forma que a esteira estivesse paralela ao eixo horizontal e recortados de forma a excluir informações externas ao conteúdo da esteira
2. Remoção de distorção óptica: OpenCV foi utilizado para remoção do efeito “olho de peixe” causado pela proximidade da esteira à câmera
3. *Deblurring*: utilizado método SNR-Deblur para remoção de embaçamento ocasionado pelo movimento da esteira

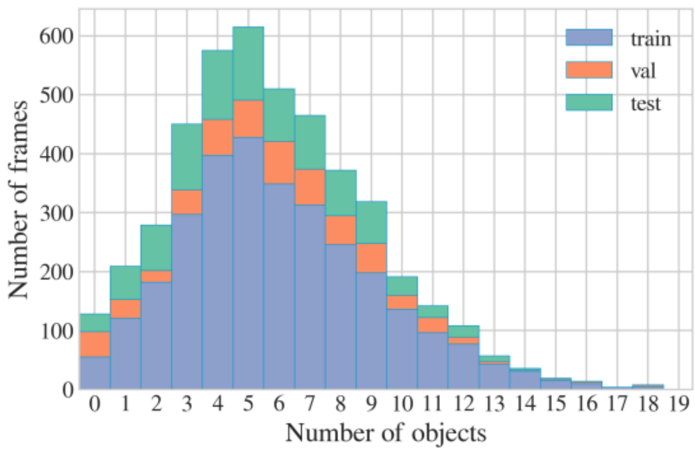
Figura 15: A esquerda – quadro original; A direita – quadro pós-processado.

Fonte: Bashkirova, 2021

Um total de 4661 quadros foram anotados de forma que apenas os contaminantes foram considerados como instâncias de interesse (*foreground*), enquanto papel e esteira foram considerados como fundo (*background*). Devido ao alto desbalanceamento entre as classes e sua inerente complexidade, optou-se por trabalhar com uma única classificação “não-papel”, tal treinamento, poderia ser útil considerando separação de materiais em uma unidade de reciclagem de papel.

O dataset anotado e armazenado sob o formato MS COCO disponibilizado pelos autores já vem dividido em conjuntos de treino, validação e teste, contando com a distribuição de objetos detectáveis por quadro indicada no histograma abaixo (figura 16).

Figura 16: distribuição de contaminantes presentes por quadro



Fonte: Bashkirova, 2021

Finalmente, na figura 17 podem ser observados exemplos de saídas esperadas para duas imagens pertencentes ao ZeroWaste-f.

Figura 17 – Apresentação de dois exemplos de imagens contidas na ZeroWaste Dataset (à esquerda) e suas respectivas segmentações e rótulos (à direita).

Fonte: BASHKIROVA, 2021.

3.2 Experimentos

Experimento consistirá em utilizar a mesma implementação do algoritmo Mask R-CNN empregada pelos autores, no entanto a CNN utilizada como backbone será alterada. No artigo de lançamento da base (BASHKIROVA, 2021) o backbone empregado era do tipo ResNet-50. Neste experimento utilizaremos a implementação com um backbone mais robusto, ResNext-101.

O modelo a ser testado foi construído em cima do framework detectron2 (WU et al., 2019). Sua construção modular permite maior flexibilidade na implementação e adaptação de algoritmos de aprendizagem profunda voltados ao processamento de imagens.

Aplicando o princípio da modularidade, será investigado como a alteração do *backbone* afetará a qualidade da saída da segmentação e classificação, através da comparação com os valores de mAP para as máscaras.

3.2.1 Configurações

O para treinamento foram utilizadas 3000 imagens da base ZeroWasteAug, as quais foram processadas com batch=2, por mil iterações. Devido ao limitado tamanho do dataset e do poder computacional disponível, para obtenção de um melhor resultado, foi utilizada transferência de aprendizagem a partir de pesos treinados na base MS-COCO, fornecidos pelo pacote detectron2. Estipulou-se uma taxa de aprendizagem de 0.0025 (configuração padrão da configuração).

Aplicando o princípio da modularidade, será investigado como a alteração do backbone afetará a qualidade da saída da segmentação e classificação, através da comparação com os valores de mAP para as máscaras.

4 RESULTADOS E ANÁLISES

Após treinamento, os resultados abaixo foram obtidos.

Tabela 4: Resultados obtidos com o Mask R-CNN (ResNext-101) e comparação com valores obtidos com configurações padrão do framework detectron2 (BASHKIROVA, 2021) para precisão média das máscaras

Métrica	ResNet-50	ResNeXt-101	Melhora
AP	22.8	**42.1**	↑ **84%**
AP50	34.9	**59.7**	↑ **71%**
AP75	24.4	**44.4**	↑ **81%**
APs	4.6	**29.1**	↑ **532%**
APm	10.6	**30.5**	↑ **187%**
APl	25.8	**44.3**	↑ **71%**

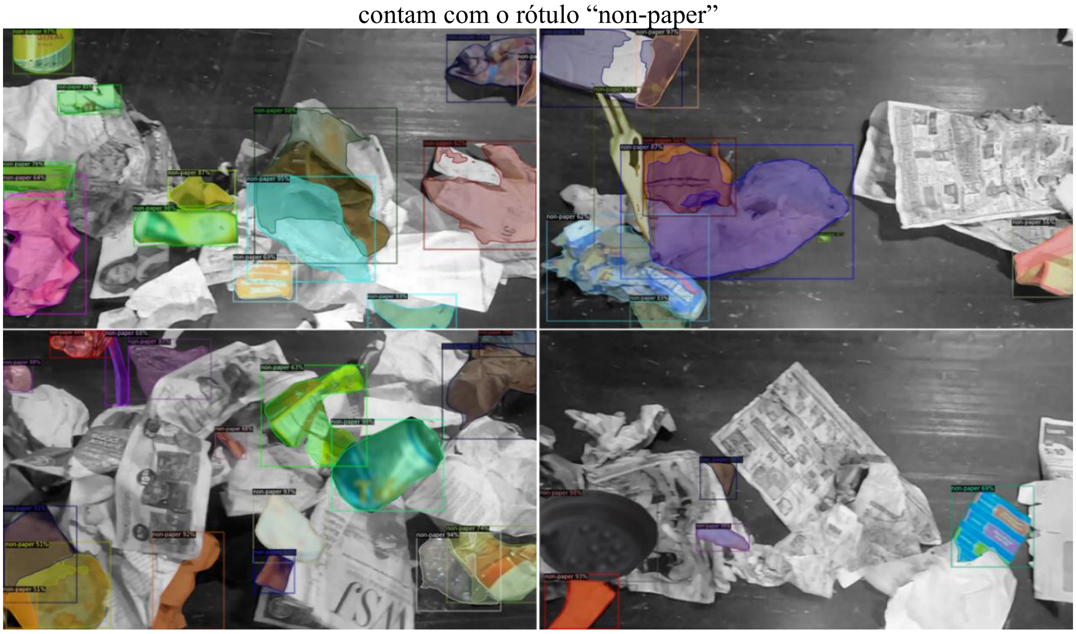
Fonte: Autoria própria

Com uma rede neural mais robusta e transferência de aprendizagem resultados a partir de 71% melhores foram obtidos. Observou-se que, assim como detectado pelos autores do dataset, o algoritmo tem dificuldade de detectar os contornos de objetos menores.

Quanto ao tempo de processamento, obteve-se 0,368 segundos/imagem para inferência na base de teste, significativamente acima dos 0,033 segundos/imagem, limite para considerar processamento em tempo real.

Abaixo, podem ser visualizados exemplos de imagens com máscaras detectadas pelo modelo aqui empregado.

Figura 18 – Exemplos de imagens obtidas a partir do modelo com ResNext-101, todas as instâncias contam com o rótulo "non-paper"



Fonte: Autoria própria

5 CONCLUSÃO

Neste trabalho foi apresentado um panorama geral sobre geração e gestão de resíduos sólidos urbanos (RSU), pontuando o desafio de aumentar a capacidade de processamento de RSU. Foi pontuada a oportunidade de sem empregar algoritmos de detecção de objetos neste setor, visando reduzir o gargalo de produção que a separação manual representa.

Foi feita um extenso levantamento bibliográfico sobre o tema de classificação e detecção de imagens e objetos. Propondo-se, a partir do entendimento da arquitetura das soluções atualmente disponíveis, uma análise de impacto sobre modificação da rede neural convolucional que gera as features maps utilizadas para alimentar o algoritmo Mask R-CNN.

Observou-se que tais redes possuem grande impacto sobre os resultados obtidos em uma mesma base de dados, levando a resultados com no mínimo 71% de melhora para o caso analisado.

No entanto, o desafio do escopo das instâncias a serem mapeadas, com alto nível de oclusão e deformação, ainda representa um desafio para obtenção de melhores detectores. De forma que ainda há bastante espaço para melhoria na tarefa de detecção e classificação de materiais recicláveis em ambientes produtivos, considerando não somente à acurácia dos modelos, mas também a velocidade de processamento para tarefa de inferência.

REFERÊNCIAS

ARAL, R. A.; KESKIN, S. R.; KAYA, M.; HACIÖMEROĞLU, M. Classification of TrashNet Dataset Based on Deep Learning Models. *In:* 2018 IEEE International Conference on Big Data. IEEE, 2018. p.2058-2062.

ASSOCIAÇÃO BRASILEIRA DE RESÍDUOS SÓLIDOS E LIMPEZA PÚBLICA (ABLP). Revista Limpeza Pública, ed. 86. 2014.

ASSOCIAÇÃO NACIONAL DE CATADORES E CATADORAS DE MATERIAIS RECICLÁVEIS (ANCAT). Anuário da Reciclagem 2021. Disponível em: <ancat.org.br > Acesso em: 17 de janeiro de 2022.

AVERSANI, O. Nova central de triagem de resíduos recicláveis é inaugurada em Santana de Parnaíba. **Giro S/A**. 15 de outubro de 2020. Disponível em: <https://www.girosa.com.br/cidade/nova-central-de-triagem-de-residuos-reciclaveis-e-inaugurada-em-santana-de-parnaiba>. Acesso em: 18 de janeiro de 2022.

BANDYOPADHYAY, H. An Introduction to Image Segmentation: Deep Learning vs. Traditional. V7Labs, 2022. Disponível em: <https://www.v7labs.com/blog/image-segmentation-guide>. Acesso em 01 de junho de 2022.

BASHKIROVA, D.; ZHU, Z.; AKL, J.; ALLADKANI, F.; HU, P.; ABLAVSKI, V.; CALLI, B.; BARGAL, S. A.; SAENKO, K. ZeroWaste dataset: Towards Automated Waste Recycling. 2021. Disponível em: <http://ai.bu.edu/zerowaste/>
Acesso em: 27 de setembro de 2021.

BIRCANOĞLU, C.; ATAY, M.; BEŞER, F.; GENÇ, Ö.; KIZRAC, A. RecycleNet: Intelligent Waste Sorting Using Deep Neural Networks. In: **2018 Innovations in Intelligent Systems and Applications (INISTA)**, 2018, p.1-7.

BRITTO, P. M. Modelos de referência de atividades operacionais aplicáveis a organizações de catadores de materiais recicláveis. Orientador: Prof. Dr. Renato Ribeiro Siman. 2018. 128 f. Dissertação (Mestrado) – Programa de Pós-Graduação em Engenharia e Desenvolvimento Sustentável, Universidade Federal do Espírito Santo, Vitória, 2018.

CHENG, T.; WANG,X; CHEN, S.; ZHANG, W.; ZHANG, Q.; HUANG, C.; ZHANG, Z.; LIU. W. Sparce Instance Activation for Real-Time Instance Segmentation. Disponível em: <https://arxiv.org/pdf/2203.12827v1.pdf>. Acesso em 26 de junho de 2022.

DATA SCIENCE ACADEMY. Deep Learning Book, 2021.
Disponível em: <https://www.deeplearningbook.com.br/>.
Acesso em: 27 de setembro de 2021

DENG, J.; DONG, W., SOCHER, R.; LI, L.; LI, K.; FEI-FEI, L. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. 2009. p. 248–55

DEWULF, V. Application of machine learning to waste management: identification and classification of recyclables. Department of Civil and Environmental Engineering. Imperial College London, 2017.

ELGENDY, M. Deep Learning for Vision Systems. Manning Publications, 1ª edição. Novembro de 2020. ISBN: 9781617286192

GOMES, L. Catadores reclamam da falta de repasse e queda nos recicláveis: ‘DMLU quer fazer o sistema morrer à míngua’. **Sul21**. 12 de junho de 2018.
Disponível em: https://sul21.com.br/cidadesz_areazero/2018/06/catadores-reclamam-de-falta-de-repasses-e-queda-nos-reciclaveis-dmlu-quer-fazer-o-sistema-morrer-a-mingua/
Acesso em: 18 de janeiro de 2022.

GUO, H. N.; WU, S. B.; TIAN, Y. J.; ZHANG. J.; LIU, H. T. Application of machine learning methods for the prediction of organic solid waste treatment and recycling processes: A review. **Bioresource Technology**, v. 319, 2021.

GUPTA, P. K.; SHREE, V.; HIREMATH, L.; RAJENDRAN, S. (2019) The Use of Modern Technology in Smart Waste Management and Recycling: Artificial Intelligence and Machine Learning. In: Kumar R., Wiil U. (eds) **Recent Advances in Computational Intelligence. Studies in Computational Intelligence**, vol 823. Springer, Cham. https://doi.org/10.1007/978-3-030-12500-4_11

HADDAD, F. R.; SILVA, D. P.; MASSOLA, C. P.; MORAES, S. L.; BERGERMAN, M. G.; Métodos de Triagem de Materiais Recicláveis: Análise Comparativa de Cooperativas do Município de São Paulo , p. 205 -214. In: **Catadores e espaços de (in)visibilidades**. São Paulo: Blucher, 2020. ISBN: 9788580394108, DOI 10.5151/9788580394108-11

HE, Y.; GU, Q.; SHI, M. Trash Classification Using Convolutional Neural Networks. CS230 Project Report, 2020.
Disponível em: <http://cs230.stanford.edu/projects_spring_2020/reports/38847029.pdf>

HUANG, G. L.; HE, J.; XU, Z.; HUANG, G. A combination model based on transfer learning for waste classification. **Concurrency Computat Pract Exper**. 2020. e5751.

HUI, J. Image segmentation with Mask R-CNN. Medium, 2018. Disponível em <https://jonathan-hui.medium.com/image-segmentation-with-mask-r-cnn-ebe6d793272>.
Acesso em 1 de junho de 2022.

KAZA, S.; YAO, L. C.; Bhada-Tata, P.; VAN WOERDEN, F. 2018. What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050. Urban Development. Washington, DC: World Bank. © World Bank. https://openknowledge.worldbank.org/handle/10986/30317 License: CC BY 3.0 IGO.

KULKARNI, H. N.; RAMAN, N. K. S. Waste Object Detection and Classification. CS230 Project Report, 2019.
Disponível em: <https://cs230.stanford.edu/projects_fall_2019/reports/26262187.pdf>. Acesso em 20 de novembro de 2021.

LEE, Y.; PARK, J. CenterMask: Real-Time Anchor-Free Instance Segmentation. CVPR, 2020. Disponível em <https://arxiv.org/abs/1911.06667>. Acesso em 26 de junho de 2022.

MAO, W. L.; CHEN, W. C.; WANG, C. T.; LIN, Y. H. Recycling waste classification using optimized convolutional neural network. **Resources, Conservation and Recycling 164: 105132**, 2020.

MAJCHROWSKA, S.; MIKOŁAJCZYK, A.; FERLIN, M.; KLAWIKOWSKA, Z.; PLANTYKOW, M. A.; KWASIGROCH, A.; MAJEK, K. Waste Detection in Pomerania: Non-Profit Project for Detecting Waste in Environment. 2021. Disponível em < https://arxiv.org/pdf/2105.06808.pdf> Acesso em: 22 de novembro de 2021.

MECHEA, D. What is panoptic segmentation and why you should care. Medium, 2019. Disponível em: <https://medium.com/@danielmechea/what-is-panoptic-segmentation-and-why-you-should-care-7f6c953d2a6a>. Acesso em 20 de junho de 2022.

MEIRA, N. Edge AI – MaskRCNN e Segmentação de Instâncias. Departamento de Computanção da Universidade Federal de Ouro Preto (UFOP), 2020. Disponível em: <http://www2.decom.ufop.br/imobilis/segmentacao-instancias/>. Acesso em 20 de junho de 2022.

OZDEMIR, M. E.; ALI, Z.; SUBESHAN, B.; ASMATULU, E. Applying machine learning approach in recycling. **Journal of Material Cycles and Waste Management.** Vol. 23, p.855-871. 2021

ÖZKAYA, U.; SEYFI, L. Fine-Tuning Models Comparisons on Garbage Classification for Recyclability**. arXiv preprint arXiv:1908.04393**, 2019.

PARREIRA, G. F.; OLIVEIRA, F. G.; LIMA, F. P. A. O gargalo da reciclagem: determinantes sistêmicos da triagem de materiais recicláveis. **Encontro Nacional De Engenharia De Produção**, 2009.

PONTI, M. A.; COSTA, G. B. P.; Como funciona o deep learning. *In*: VIEIRA, V.; RAZENTE, H. L.; BARIONI, M. C. N. **Tópicos em Gerenciamento de Dados e Informações**. Minas Gerais: Sociedade Brasileira de Computação, 2017. p.63-93

PONTI, M. A. (3) Introdução ao Aprendizado Profundo. *In:* Redes Neurais e Deep Learning – MBA de Inteligência Artificial e Big Data. ICMC-USP, 2021.

RAMSURRUN, N.; SUDDUL, G.; ARMOOGUM, S.; FOOGOOA, R. Recyclable Waste Classification Using Computer Vision and Deep Learning. *In:* **Zooming Innovation in Consumer Technologies Conference (ZINC)**, 2021, p.11-15.

REDE NOSSA SAO PAULO (RNSP). Cidade ganha segunda central mecanizada de triagem de resíduos sólidos. **RNSP**. 18 de julho de 2014. Disponível em: < https://www.nossasaopaulo.org.br/2014/07/18/cidade-ganha-segunda-central-mecanizada-de-triagem-de-residuos-solidos/>. Acesso em: 18 de janeiro de 2022.

SALMI, J. Developing Computer Vision-based soft sensor for municipal solid waste burning grate boiler – A Practical Application for Flame Front and Area Detection. Master of Science

Thesis. Tampere University. Master’s Degree Program in Automation Engineering. Setembro de 2021. Disponível em: <https://trepo.tuni.fi/bitstream/handle/10024/134257/SalmiJesse.pdf?sequence=2>. Acesso em 20 de junho de 2022.

SEREDKIN, A. V.; TOKAREV, M. P.; PLOHIH, I. A.; GOBYZOV, O. A.; MARKOVICH, D. M. Development of a method of detection and classification of waste objects on a conveyor for a robotic sorting system. Journal of Physics: Conference Series 1359 012127. 2019.

SIDHART, R.; ROHIT, P.; VISHGAN, S.; KARTHIKA, S.; GANESAN, M. Deep Learning based Smart Garbage Classifier for Effective Waste Management. *In:* **2020 5th International Conference on Communication and Eletronics Systems (ICCES)**. IEEE, 2020. p.1086-1089.

SINGH, S.; LUO, M.; LI, Y. Generalized Anomaly Detection. Portland State University, 2021. Disponível em < https://arxiv.org/pdf/2110.15108.pdf>. Acesso em: 22 de novembro de 2021.

V7 LABS. The Definitive Guide to Instance Segmentation, 2022. Disponível em < https://www.v7labs.com/blog/instance-segmentation-guide >. Acesso em 20 de junho de 2022.

VO, A. H.; SON, L. H.; VO; M. T.; LE, T. A Novel Framework for Trash Classification Using Deep Transfer Learning. *In:* IEEE Access, vol. 7, p.178631-178639, 2019.

WU, Y.; KIRILLOV, A.; MASSA, F.; LO, W.; GIRSHIK, R. Detectron2. Disponível em <https://github.com/facebookresearch/detectron2>. Acesso em 02 de julho de 2022.

YANG, M.; THUNG, G. Dataset of images of trash; Torch-based CNN for garbage image classification. GitHub Repository, 2016.
Disponível em: <https://github.com/garythung/trashnet>
Acesso em: 29 de setembro de 2021.