

INTELIGÊNCIA ARTIFICIAL E SENSORIAMENTO REMOTO COM O USO DE VANTS: UMA APLICAÇÃO NA DETECÇÃO AUTOMÁTICA DE RESÍDUOS SÓLIDOS URBANOS

Data de submissão: 08/11/2024

Data de aceite: 02/12/2024

Marcelo Musci

Universidade do Estado do Rio de Janeiro
– UERJ
<https://orcid.org/0000-0001-8728-9078>

Carlos Vitor de Alencar Carvalho

Universidade do Estado do Rio de Janeiro
– UERJ
<https://orcid.org/0000-0003-1078-3689>

Gabriel de Mello Pereira Serrão

Universidade do Estado do Rio de Janeiro
– UERJ
<https://orcid.org/0009-0001-4363-3225>

Marcos Vinícius Elias Neres Barreto Ferreira

Universidade do Estado do Rio de Janeiro
– UERJ
<https://orcid.org/0009-0005-6793-0020>

Maycow Duarte Pinto Guerra

Universidade do Estado do Rio de Janeiro
– UERJ
<https://orcid.org/0009-0000-5432-077X>

Flavio Lucas dos Santos Baptista

Universidade do Estado do Rio de Janeiro
– UERJ
<https://orcid.org/0009-0005-4576-8691>

Giancarlo Cordeiro da Costa

Universidade do Estado do Rio de Janeiro
– UERJ
<https://orcid.org/0009-0005-4576-8691>

RESUMO: De acordo com o Instituto de Pesquisa Econômica Aplicada (IPEA), aproximadamente 80 milhões de toneladas de resíduos sólidos urbanos foram geradas no Brasil em 2018, e uma parte significativa desses resíduos não recebeu o descarte adequado, sendo irregularmente descartada em áreas urbanas ou periféricas, encostas e locais de difícil acesso. Esse problema é ainda mais pronunciado em regiões onde o Estado falha devido à falta de presença e assistência, fazendo com que os moradores recorram ao descarte irregular de resíduos em regiões montanhosas e de difícil acesso. Na verdade, mesmo em áreas mais desenvolvidas, o acúmulo de tais resíduos pode levar a deslizamentos de terra e à proliferação de ratos, baratas, moscas e outros vetores urbanos. Dada a complexidade de identificar locais de descarte irregular de resíduos em áreas remotas, seja por omissão do governo ou pelo difícil acesso a esses locais, este

estudo tem como objetivo apresentar os veículos aéreos não tripulados (VANTs) como uma alternativa para o monitoramento ambiental. Esses dispositivos possuem várias vantagens, como mobilidade e aquisição de imagens em tempo real de locais remotos e de difícil acesso, com custos mais baixos e sem riscos para o operador. No entanto, devido ao terreno acidentado e à vasta extensão territorial do Rio de Janeiro, o volume de dados coletados por esses dispositivos exigiria um esforço significativo para análise e interpretação manuais. Para resolver essa limitação, acreditamos que as imagens e vídeos coletados nesta pesquisa devem ser interpretados, georreferenciados e segmentados automaticamente utilizando Inteligência Artificial. Assim, com uma abordagem combinada que consiste no uso de um dispositivo eletrônico para sensoriamento remoto (VANT) equipado com câmeras de alta precisão, sensores e um sistema de posicionamento global (GPS), juntamente com recursos computacionais para o processamento e interpretação de imagens, este projeto teve como objetivo demonstrar a viabilidade dessa técnica na identificação precoce e eficaz de áreas remotas com descarte irregular de resíduos. Após um processo completo de sensoriamento remoto realizado com o auxílio de um VANT nas áreas de estudo deste projeto, as imagens capturadas foram processadas em lote utilizando técnicas de processamento digital de imagens e algoritmos de Deep Learning. Após um extenso treinamento, esses algoritmos foram capazes de segregar e rotular automaticamente as imagens que continham evidências de descarte de resíduos sólidos daquelas que não continham. Como resultado dessa segregação automática, a técnica demonstrou uma eficácia de 92% na identificação de tipos específicos de resíduos, ajudando assim a identificar automaticamente áreas carentes que dependem da ação governamental para combater e mitigar o descarte irregular de resíduos e seu impacto ambiental.

PALAVRAS-CHAVE: Sensoriamento remoto, VANTs, Inteligência Artificial, Detecção automática de Resíduos sólidos

ABSTRACT: According to the Institute of Applied Economic Research (IPEA), approximately 80 million tons of solid urban waste were generated in Brazil in 2018, a significant portion of which did not receive proper disposal and was irregularly discarded in urban or peripheral areas, slopes, and hard-to-reach areas. This problem is even more pronounced in regions where the state fails due to lack of presence and assistance, resulting in residents resorting to irregular waste disposal in hilly and hard-to-reach regions. In fact, even in more developed areas, the accumulation of such waste can lead to landslides and the proliferation of rats, cockroaches, flies, and other urban pests. Given the complexity of identifying irregular waste disposal sites in remote areas, either due to government omission or difficult access to these locations, this study aims to present unmanned aerial vehicles (UAVs) as an alternative for environmental monitoring. These devices have several advantages, such as mobility and real-time image acquisition from remote and hard-to-reach locations, with lower costs and no risks to the operator. However, due to the rugged terrain and vast territorial expanse of Rio de Janeiro, the volume of data collected by these devices would require significant effort for manual analysis and interpretation. To address this limitation, we believe that the images and videos collected in this survey should be automatically interpreted, georeferenced, and segmented using Artificial Intelligence. Thus, with a combined approach consisting of using an electronic device for remote sensing (UAV) equipped with high-precision cameras,

sensors, and a global positioning system (GPS), along with computational resources for image processing and interpretation, this project aimed to demonstrate the feasibility of this technique in the early and effective identification of remote areas with irregular waste disposal. After a thorough remote sensing process carried out with the assistance of a UAV in the study areas of this project, the captured images were batch processed using digital image processing techniques and Deep Learning algorithms. After extensive training, these algorithms were able to automatically segregate and label images that contained evidence of solid waste disposal from those that did not. As a result of this automatic segregation, the technique demonstrated an 92% effectiveness in identifying specific types of waste, thus helping to automatically identify underserved areas that depend on government action to curb and mitigate irregular waste disposal and its environmental impact.

KEYWORDS: Remote sensing, UAVs, Automatic Waste Detection, Artificial Intelligence.

INTRODUÇÃO

O descarte inadequado de resíduos está se tornando uma preocupação global, especialmente em áreas metropolitanas com populações densas e escasso apoio governamental. Consequências significativas para a saúde pública e a deterioração ambiental podem resultar de coleta, tratamento e descarte final inadequados de resíduos sólidos (Antenor; Szigethy, 2020). Ferramentas de monitoramento têm sido amplamente utilizadas para monitorar e prevenir o descarte inadequado de lixo, a fim de apoiar a resolução deste problema (Glanville; Chang, 2015).

O monitoramento ambiental pode se beneficiar do uso de satélites e drones de alta resolução. Os drones são mais apropriados para áreas menores, pois conseguem capturar imagens com um nível de detalhe maior do que os satélites, que são destinados ao monitoramento de áreas mais amplas. O monitoramento de resíduos tem sido aprimorado globalmente por esses dispositivos potenciais, embora ainda enfrente limitações no setor público devido ao financiamento para tecnologia e conformidade regulatória. Satélites e VANTs (Veículos Aéreos não Tripulados), também conhecidos como drones, são duas tecnologias que podem ser utilizadas para o monitoramento ambiental (Alvarez-Vanhard, 2021).

Os drones podem tirar fotos excelentes em espaços limitados, mas não são capazes de cobrir áreas muito grandes. No entanto, eles podem capturar detalhes finos de objetos e elementos ambientais com uma resolução espacial muito maior do que os satélites, quando equipados com câmeras e sensores de alta resolução (Green et al., 2017). Eles também podem tirar fotos de diferentes ângulos e a altitudes mais baixas, oferecendo uma perspectiva ainda mais precisa e detalhada de determinadas regiões. Comparado com outras técnicas de monitoramento, o uso de drones para monitoramento ambiental oferece algumas vantagens, especialmente no que diz respeito ao descarte inadequado de resíduos.

Além de serem capazes de capturar imagens precisas de colinas e áreas de difícil acesso sem serem impactados por limitações geográficas, o uso de drones permite maior precisão e rapidez na georreferência de áreas suspeitas de descarte inadequado de lixo (Youme et al., 2021). Com o uso de todos os dados coletados pelos sensores, é possível criar mapas abrangentes das práticas inadequadas de descarte de lixo na cidade, e programar ações de preservação ambiental e mitigação. A operação em tempo real é outro benefício do uso de drones no monitoramento ambiental. As informações obtidas podem ser usadas para monitorar a eficácia das regulamentações de gestão de resíduos e indicar áreas que necessitam de melhorias.

O objetivo deste trabalho é mostrar como o lixo pode ser automaticamente detectado em fotografias tiradas por drones como parte de operações de sensoriamento remoto, utilizando inteligência artificial. Nesse contexto, vários pesquisadores têm utilizado essa tecnologia para monitorar e capturar imagens, que depois são analisadas e interpretadas para o monitoramento de resíduos plásticos em regiões costeiras, como a criação de um método para rastrear lixo flutuante em corpos d'água e um mapa de locais de despejo ilegal de resíduos (Glanville; Chang, 2015; Geraeds et al., 2019).

REFERENCIAL TEÓRICO

Drones no monitoramento ambiental

Veículos aéreos não tripulados (VANTs), ou drones, são aeronaves controladas remotamente. Suas aplicações variam desde o monitoramento de projetos de construção civil até a necessidade de registros fotográficos de alta qualidade. Posteriormente, seu potencial para o monitoramento ambiental foi reconhecido, e organizações governamentais e empresas privadas começaram a utilizar esses dispositivos para realizar tarefas ambientais, como avaliação de impactos, prevenção de desastres naturais e proteção da fauna e flora em diferentes biomas (Manfreda et al., 2018).

Deep learning

As redes neurais, que são modeladas seguindo a estrutura e o funcionamento dos sistemas cerebrais orgânicos, são essenciais para a aplicação do aprendizado profundo. Essas redes, compostas por neurônios artificiais interconectados, processam e modificam os dados recebidos por meio de camadas de processamento (Ren et al., 2016). Além de realizar tarefas como regressão, classificação e processamento de dados não lineares, as redes neurais podem aprender representações de dados complexos. O algoritmo de retropropagação e outras técnicas de aprendizado são usadas para treinar redes neurais, ajustando os pesos e vieses das conexões entre os “neurônios”. As redes neurais possuem várias formas, cada uma com características e aplicações específicas, incluindo: Redes

Neurais Feedforward (FFNN), Redes Neurais Recorrentes (RNN), e Redes Neurais Convolucionais (CNN).

As FFNNs são compostas por camadas de neurônios com um único caminho de fluxo de dados, desde a camada de entrada até a camada de saída. Elas são frequentemente utilizadas em tarefas como regressão e classificação. Ao contrário das FFNN, as RNNs incluem conexões que formam ciclos, permitindo que os dados sejam enviados de volta para camadas anteriores. Elas são adequadas para aplicações baseadas em dados sequenciais, como processamento de linguagem natural e análise de séries temporais. Além disso, a criação das CNNs foi focada na análise de dados em formato de grade, como imagens. Elas possuem camadas convolucionais que aplicam filtros para extrair características relevantes das imagens e camadas de pooling que reduzem a dimensionalidade. As CNNs são frequentemente usadas em aplicações de visão computacional, como classificação de imagens e reconhecimento de objetos (Li; Zhang, 2020).

YOLO

Uma das ferramentas de visão computacional que tem ganhado bastante atenção nos últimos anos é uma rede neural convolucional (CNN) chamada YOLO: *You Only Look Once* (Você Olha Apenas Uma Vez). Quando o YOLO foi lançado pela primeira vez em 2015, foi imediatamente reconhecido como uma técnica inovadora, pois utilizava uma metodologia completamente nova para alcançar uma precisão equivalente ou superior às técnicas anteriores de reconhecimento de objetos da época, tudo isso enquanto detectava objetos de forma significativamente mais rápida (Xu; Wu, 2020).

Áreas de alta pontuação da imagem eram consideradas como detecções em sistemas de detecção de objetos anteriores ao YOLO, que utilizavam classificadores para realizar a identificação aplicando um algoritmo à imagem em múltiplos locais e tamanhos. *Bounding Boxes* (caixas delimitadoras) são as pequenas seções extraídas da imagem, e cada uma delas detecta um único objeto. Para classificar objetos, esse método de classificação examina a imagem em diferentes tamanhos e localizações. O YOLO, por sua vez, precisa olhar para a imagem apenas uma vez antes de enviá-la para a rede neural, em contraste com algoritmos que escaneiam a imagem em busca de itens aprendidos. Para isso, a rede neural divide a imagem em seções menores e faz previsões sobre as caixas delimitadoras, a probabilidade de um item estar dentro daquela caixa e a probabilidade de o objeto ser um determinado tipo de objeto (Redmon; Farhadi, 2017).

Com o mesmo nome do framework utilizado para implementá-lo, o YOLO emprega uma rede neural profunda (DNN), uma rede neural convolucional, cuja arquitetura é chamada Darknet. Embora a comunidade e diversas organizações tenham contribuído para seu desenvolvimento, sua implementação inicial foi criada na linguagem de programação C e agora está disponível em várias outras linguagens de programação. Esse tipo de

abordagem é utilizado para segmentar automaticamente fotos, como mostrado na Figura 1.



Figura 1 - Segmentação automática de dados usando rede YOLO

Fonte: Mittal *et al*, 2016.

O retângulo que contém a localização do objeto é chamado de *bounding box* (caixa delimitadora). Durante o processo de detecção de objetos, podem ser encontradas múltiplas caixas delimitadoras para o mesmo objeto. Essas caixas também podem ter tamanhos diferentes. Cada caixa tem uma pontuação de confiança, que indica o nível de certeza que o algoritmo tem de que a caixa contém um objeto.

A quarta versão do YOLO foi lançada em abril de 2020, sendo oficialmente introduzida após a publicação do artigo “YOLOv4: Optimal Speed and Accuracy of Object Detection” (Bochkovskiy, Wang & Liao, 2020).

As principais características que podem ser destacadas nesta versão são a melhoria na velocidade de inferência e precisão. Outra característica importante é sua eficiência, pois foi otimizada para utilizar menos memória. O YOLOv4 provou ser o melhor detector de objetos para testes em tempo real, de acordo com as métricas do dataset MS COCO (Tsung-Yi, 2014), um famoso conjunto de dados utilizado para avaliar sistemas de detecção de objetos.

Na Figura 2, as caixas delimitadoras detectadas pelo algoritmo YOLO estão destacadas. A pontuação de confiança também está destacada na imagem, o que significa que, quanto maior a confiança do YOLO de que um objeto existe dentro daquela caixa, maior será a pontuação da caixa. O YOLO cria múltiplas caixas delimitadoras. Para cada caixa, um valor de confiança é atribuído, indicando a porcentagem (de 0 a 1) da existência do objeto.



Figura 2 - Caixas delimitadoras de objetos detectados pela YOLO e seu respectivo valor de confiança

Fonte: Elaborado pelos Autores (2023).

O processo de treinamento dos algoritmos pode envolver várias etapas, dependendo de suas características. Nos algoritmos supervisionados, o algoritmo é treinado utilizando um grande conjunto de imagens rotuladas, ou seja, imagens que foram previamente classificadas como áreas limpas ou áreas com a presença de resíduos. Essas imagens de treinamento fornecem informações supervisionadas para que o algoritmo aprenda a reconhecer padrões específicos de descarte irregular de resíduos. Esse processo pode ser detalhado da seguinte forma:

- Coletar um grande conjunto de imagens, que serão posteriormente rotuladas, para serem usadas como dados de treinamento;
- Pré-processar as imagens, como redução de ruído, ajuste de contraste e brilho, e normalização de pixels;
- Treinar o algoritmo utilizando técnicas de aprendizado supervisionado para reconhecer padrões em imagens segmentadas que foram previamente rotuladas como contendo exemplos de descarte de resíduos e outras livres desse problema.
- Ao término, sua eficácia é medida para avaliar a precisão da detecção automática.
- O algoritmo pode passar por calibrações sucessivas até que a eficácia esperada seja atingida.

Ao final do treinamento, como mencionado por Verma et al. (2022), os algoritmos devem ser capazes de identificar automaticamente a presença de resíduos em imagens capturadas pelo drone. Isso permitirá que os dados coletados sejam processados e analisados de forma mais eficiente, ajudando a identificar áreas de risco e planejar ações de limpeza e prevenção para o descarte irregular, mesmo em grandes volumes de dados.

Durante o treinamento de uma rede neural, o valor esperado para cada objeto a ser detectado, em uma iteração, é comparado com os valores de saída. Cada ciclo de treinamento é chamado de época (epoch), e quanto maior o número de épocas, mais fácil será para a rede trabalhar com os valores de treinamento. É importante lembrar que tanto o treinamento insuficiente quanto o treinamento excessivo podem ter um impacto negativo na rede. Com treinamento insuficiente, a rede não consegue aprender a lidar com os valores de treinamento, o que é um problema chamado *underfitting*. Com treinamento excessivo, a rede pode ficar condicionada aos valores de treinamento e pode não ter sucesso ao receber valores diferentes, o que é chamado de *overfitting*. Além disso, durante o treinamento, o valor chamado *loss* (perda) representa a diferença entre o valor de saída da rede e o valor esperado para uma entrada dada. A diminuição gradual do valor de perda a cada época indica que a rede está fazendo previsões mais bem-sucedidas (Filipakis; Fagundes, 2021).

METODOLOGIA

Este trabalho foi realizado em várias etapas, sendo que cada uma delas utilizou uma metodologia específica que se adapta aos diferentes insumos, ferramentas e técnicas necessárias em cada fase. No final, os resultados obtidos em cada etapa serão combinados para criar um projeto que integra diversos campos do conhecimento e analisa todos os aspectos relacionados ao uso de drones para monitoramento ambiental e ao processamento automático das imagens capturadas por esses dispositivos.

Mapeamento da área de estudo

Para avaliar os impactos ambientais, é importante definir o espaço geográfico e compreender as informações relacionadas ao uso e ocupação do solo, topografia e outros dados relevantes. Os drones auxiliam as equipes que precisam realizar o monitoramento ambiental, permitindo analisar a área de estudo de forma mais concisa e eficiente, fornecendo imagens georreferenciadas, além de informações sobre posição e altitude.

Esses estudos ambientais estão intimamente ligados à geografia do objeto de estudo, pois a localização e outros aspectos geográficos são essenciais para verificar a precisão do levantamento, especialmente aqueles que são o foco desta pesquisa, como a captura de imagens contendo evidências de descarte de resíduos em encostas e áreas de difícil acesso. Por isso, os drones são tão úteis na avaliação, pois permitem o levantamento e a quantificação das características relevantes para esta pesquisa sem expor a equipe do projeto a riscos.

O equipamento utilizado, o drone Phantom 4 Multispectral, possui seis sensores CMOS de 1/2.9", incluindo um sensor RGB para imagens em luz visível e cinco sensores monocromáticos para imagens multiespectrais, com autonomia de voo de até 20 minutos. Ele está equipado com uma câmera capaz de gravar vídeos com resolução de até 1080×720

pixels a 60 Hz. Os vídeos gerados foram gravados em 1080p a 30 quadros por segundo (fps).

Para treinar a rede YOLOv4, utilizada neste trabalho, foi necessário construir uma base de dados contendo imagens. Para isso, foi realizado um levantamento de locais contendo resíduos sólidos (lixo), e cinco vídeos foram obtidos, compondo assim a base de dados de treinamento e teste.

O conjunto de dados possui as seguintes especificações técnicas, incluindo:

- Um plano de voo foi definido usando o software DJI GS PRO. Este plano realiza um voo em serpentina por toda a área do terreno de forma autônoma. O ajuste automático da câmera foi desativado, e todos os parâmetros foram configurados manualmente, mantendo o foco fixo no infinito e ajustando a resolução do vídeo para 3840 pixels a 50 quadros por segundo (fps).

- A altitude é de aproximadamente 10 metros e permanece constante em cada vídeo. Pequenas variações na altitude não influenciam a construção da base de dados e estão dentro de faixas aceitáveis.
- O conjunto de dados inclui diferentes tipos de terreno, todos contendo vários tipos de resíduos sólidos (lixo).
- O conjunto de dados gerado contém múltiplas sequências de vídeos aéreos, dos quais foram extraídas imagens e manualmente anotadas com os objetos de interesse (lixo), utilizando *bounding boxes*.

As amostras (imagens) dos cinco vídeos capturados em diferentes locais, listados na Tabela 1, podem ser visualizadas na Figura 3.

A base de dados atualmente consiste em 5 sequências de vídeos, adquiridos por um drone, gravados em 5 locais diferentes, conforme detalhado na Tabela 1. Todos os vídeos foram gravados no Rio de Janeiro, Brasil. A base de dados contém 4.169 imagens com objetos de interesse anotados. O grande número de *bounding boxes* anotadas possibilitou o treinamento de detectores de objetos usando a rede YOLOv4.

Vídeo	LATITUDE	LONGITUDE
1	- 22.896806°	- 43.584423°
2	- 22.935590°	- 43.700362°
3	- 22.926178°	- 43.694735°
4	- 22.862512°	- 43.531212°
5	- 22.922022°	- 43.635188°

Tabela 1 – Localização das áreas mapeadas para construção da base de dados - Valores de latitude e longitude da cidade do Rio de Janeiro

Fonte: Google Maps (n.d.). [Rio de Janeiro].

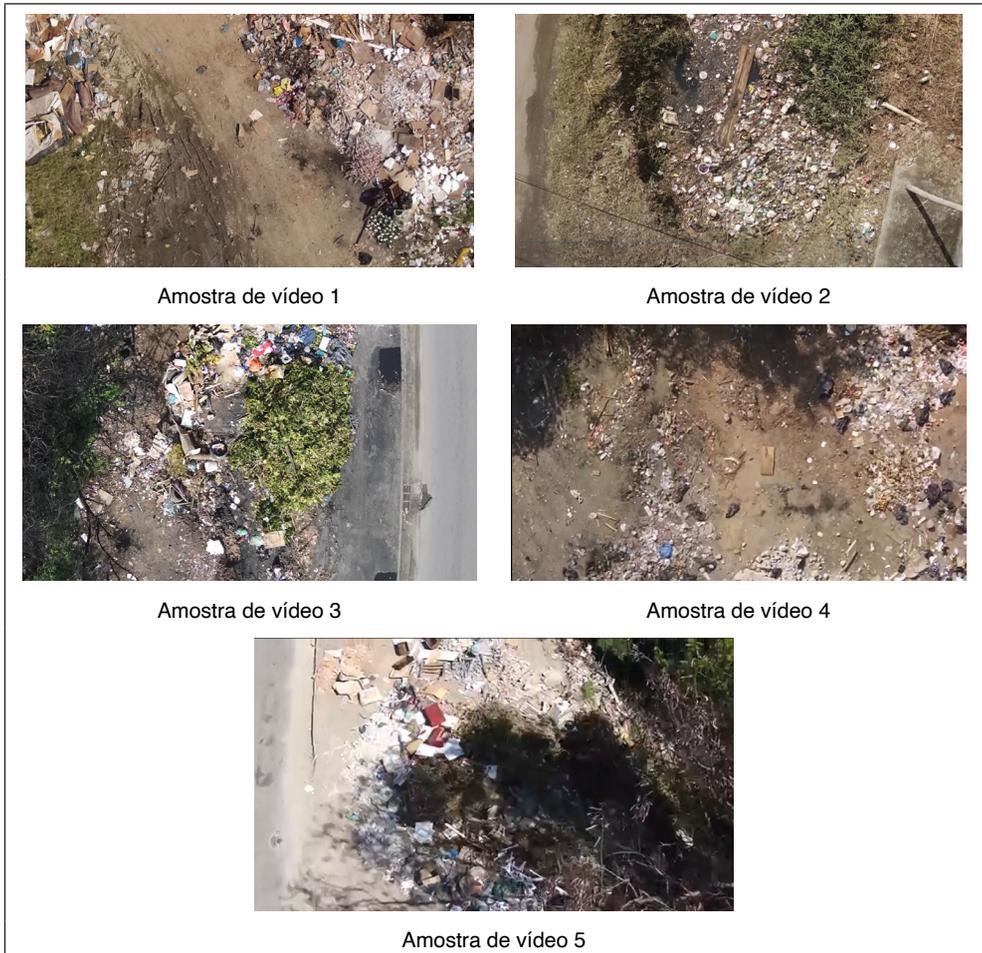


Figura 3 - Imagens dos vídeos capturados em diferentes locais listados na Tabela 1.

Fonte: Elaborado pelos Autores (2023).

Deep learnig

Um grande avanço no monitoramento ambiental é a aplicação do aprendizado profundo em sensoriamento remoto. Essa técnica melhora a interpretação das fotos obtidas por drones e satélites, pois essas imagens podem fornecer informações importantes sobre o ambiente, incluindo a quantidade de vegetação, a qualidade da água e a presença de contaminantes (Youme et al., 2021). No entanto, devido à grande quantidade e complexidade dessas imagens, avaliá-las pode ser desafiador. Por isso, o aprendizado profundo é utilizado nesta situação. Redes neurais convolucionais (CNNs) são um tipo de rede neural profunda que pode extrair automaticamente informações das imagens, como texturas, formas e padrões, proporcionando uma nova perspectiva para entender e lidar com os desafios ambientais atuais e futuros.

Para o desenvolvimento do sistema proposto de detecção de resíduos sólidos, a base de dados foi inicialmente dividida em conjuntos de treinamento e teste. Os vídeos 2, 3, 4 e 5 foram usados para treinamento, e o vídeo 1 foi utilizado para testar o sistema proposto, utilizando a rede YOLOv4.

Uma característica da rede YOLOv4 é que apenas imagens contendo os objetos a serem detectados (resíduos) devem ser usadas em seu treinamento. Portanto, todas as imagens contidas nos vídeos de treinamento devem ser cuidadosamente selecionadas manualmente para garantir que os objetos de interesse estejam presentes na cena. Após o treinamento, a rede estará pronta para detectar objetos presentes no vídeo de teste (vídeo 1).

Para o treinamento da rede YOLOv4, foi utilizado um computador de alto desempenho, um servidor Intel i9 com 12 núcleos de processamento, 128 GB de memória e uma placa de vídeo Nvidia GeForce RTX2080/11GB. Os algoritmos de treinamento utilizam as seguintes bibliotecas de código:

- CUDA e cuDNN: Desenvolvidas pela Nvidia, essas bibliotecas permitem a execução de algoritmos de alto custo computacional em dispositivos GPU.
- Darknet: Contém implementações para as quatro primeiras versões da arquitetura YOLO, sendo esse framework desenvolvido pelos autores da rede.
- LabelImg: Um software de código aberto desenvolvido em Python, usado para anotar imagens do conjunto de dados e exportar seus respectivos arquivos de anotação para vários formatos de treinamento.
- Pytorch: Fornecido pela equipe de inteligência artificial do Facebook, contém a implementação da rede YOLO para sua quarta versão.
- Roboflow: Uma plataforma online focada em problemas de visão computacional, desde pré-processamento e anotação até as etapas de treinamento de modelos.
- TensorFlow 2: Uma biblioteca de código aberto desenvolvida pelo Google, onde é possível configurar e executar treinamentos para diferentes arquiteturas de redes neurais, além de uma grande variedade de tarefas em Inteligência Artificial.

ANÁLISE DE RESULTADOS

Utilizando os conjuntos de dados para treinamento e teste, descritos na seção anterior, o modelo de rede neural YOLOv4 foi projetado para o experimento neste estudo. O modelo de rede possui apenas duas classes, classe-0 e classe-1, respectivamente com e sem detecção de resíduos sólidos.

A rede foi treinada com 4.169 imagens rotuladas, 2000 *epochs* (iterações) e uma apresentando uma *loss* média de 4.2, conforme apresentado na Figura 4. O tempo médio

de treinamento ficou em aproximadamente 12h.

A Tabela 2 ilustra a taxa de acerto, eficácia, da rede com o vídeo 1 usado como teste, onde objetos são a presença (classe-1) ou ausência de resíduos sólidos (classe-0) nas imagens. De acordo com os resultados apresentados na Tabela 2, a rede treinada obteve uma eficácia de aproximadamente 80% para o conjunto de teste.

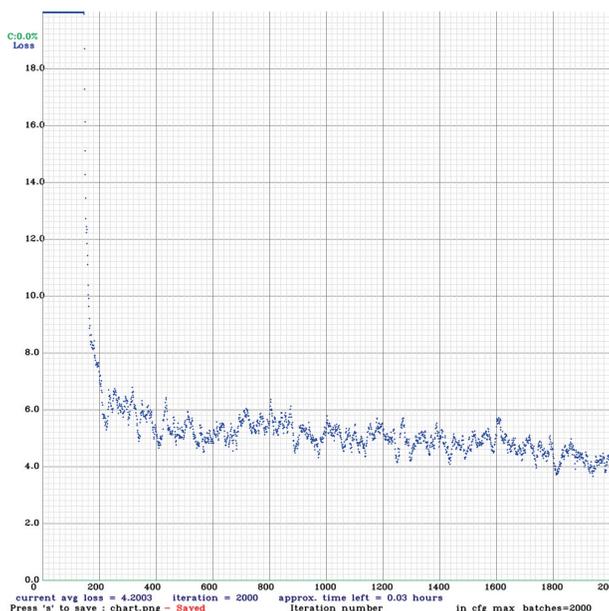


Figura 4 – Treinamento da YOLOv4, número de iterações x *loss*.

Fonte: Elaborado pelos Autores (2023).

Vídeo 1		
Nº objetos detectados	Nº objetos não-detectados	Total objetos
3.231	938	4.169

Tabela – 2 – Eficácia da rede YOLO no conjunto de teste.

Fonte: Elaborado pelos Autores (2023).

Os resultados indicam uma *loss* não ideal para o conjunto de treinamento utilizado, sendo a *loss* ideal a que mais se aproxima de zero, o que indica a necessidade de mais objetos de treinamento obtidos dos vídeos.

Análise Visual

Alguns resultados visuais foram analisados nas detecções das imagens. Tentamos associar a visualização com os resultados numéricos mostrados pela rede YOLOv4, que delimita os objetos detectados em caixas e ilustra a taxa de acerto do objeto pertencer a classe treinada.

Na Figura 5 e Figura 6 podemos observar que para diferentes frames adjacentes de vídeo (imagens), a rede não detecta de imediato o objeto (Figura 5), porém o objeto de interesse é detectado no frame subsequente (Figura 6). Ambos possuem taxa de acerto maiores que 50%.



Figura 5 – Detecção de apenas um objeto na imagem

Fonte: Elaborado pelos Autores (2023).



Figura 6 – Detecção correta de dois objetos na imagem

Fonte: Elaborado pelos Autores (2023).

Na Figura 7 e Figura 8 podemos observar o mesmo efeito de não detecção em objetos em frames de vídeo subsequentes, porém existe o agravante do objeto estar cortado no início da imagem mostrada na Figura 7.



Figura 7 – Detecção de apenas um objeto na imagem

Fonte: Elaborado pelos Autores (2023).



Figura 8 – Detecção correta de dois objetos na imagem

Fonte: Elaborado pelos Autores (2023).

A Figura 9 ilustra a detecção de um grande aglomerado de objetos, porém o ideal seria uma composição de objetos adjacentes, formando um mosaico mais fácil de ser detectado. Nesta figura podemos observar o efeito de uma marcação (*bounding box*) feita de forma incorreta.



Figura 9 – Marcação (*bounding box*) feita de forma incorreta.

Fonte: Elaborado pelos Autores (2023).

De forma geral, o sistema, em sua primeira versão, apresentou uma boa eficácia, da ordem de 80%, nas imagens apresentadas no vídeo de teste.

É importante observar que há a necessidade de objetos de treinamento adicionais, que podem ser obtidos a partir dos filmes, pois os resultados mostram que o conjunto de treinamento utilizado não alcançou o custo esperado, o qual deveria ser mais próximo de zero.

Um novo treinamento foi realizado com a utilização do mesmo número de imagens etiquetadas (4.169), porém com a realização de recortes menores nas *bounding boxes* dos objetos. Após 2000 épocas (iterações), o número de perdas (loss) foi de 2,6, alcançando resultados melhores do que os obtidos anteriormente, conforme mostrado na Tabela 3.

Vídeo 1		
Detected objects	Undetected objects	Total objects
3.831	338	4.169

Tabela 3 - Eficácia da rede YOLO no conjunto de teste (2ª versão).

Fonte: Elaborado pelos Autores (2023).

O desempenho geral do sistema nas imagens do vídeo de teste obteve uma melhora de 15% em relação ao resultado anterior, alcançando 92% de eficácia. De fato, mais imagens de treinamento, abordagens de redundância aplicadas aos quadros subsequentes das imagens e melhorias no sistema podem levar a uma eficácia ainda maior.

CONSIDERAÇÕES FINAIS

Este artigo discutiu o desafio de detectar automaticamente resíduos sólidos (lixo) e seu possível uso em áreas complexas, aplicando técnicas de visão computacional. Isso inclui um estudo de investigações anteriores com propósitos semelhantes que utilizaram procedimentos de aprendizado de máquina. Após a conclusão deste estudo, foi determinado que novos dados eram necessários para treinar modelos capazes de identificar automaticamente locais semelhantes que contivessem resíduos sólidos. Um drone foi utilizado para gravar imagens dos objetos que precisam ser identificados em uma grande região geográfica para coletar os dados. O software de mapeamento foi aplicado, e as configurações da câmera foram modificadas manualmente antes da captura. Os vídeos foram anotados manualmente com *bounding boxes* após a aquisição, o que possibilitou o treinamento e a avaliação das imagens obtidas. Um modelo de rede neural convolucional (CNN) foi investigado, considerando pesquisas anteriores na área de visão computacional para a identificação de resíduos sólidos, oferecendo uma nova perspectiva sobre este problema. A eficácia da rede neural YOLOv4 para o reconhecimento de resíduos sólidos em fotos aéreas foi analisada pelos autores neste trabalho. Em uma primeira tentativa a rede conseguiu detectar as imagens coletadas com cerca de 80% de precisão. Após modificar o tamanho das *bounding boxes* utilizadas anteriormente, de forma a capturar mais detalhes dos resíduos sólidos, foi percebido uma melhora de 15% na eficácia global da metodologia, alcançando 92%. Métodos de aprendizado profundo usando CNNs, como o YOLOv4, são um campo de pesquisa cada vez mais popular, com grandes avanços em visão computacional. A identificação automática de resíduos baseada em inteligência artificial pode ser uma alternativa viável para monitorar locais de difícil acesso. Este estudo mostra como as redes neurais podem ser usadas para aumentar a eficiência da gestão de resíduos sólidos, o que beneficiará nossa sociedade tanto ecologicamente quanto economicamente.

REFERÊNCIAS

- Alvarez-Vanhard, E., Corpetti, T., & Houet, T. (2021). UAV & satellite synergies for optical remote sensing applications: A literature review. *Science of Remote Sensing*, 3. <https://doi.org/10.1016/j.srs.2021.100019>
- Antenor, S., & Szigethy, L. (2020). Resíduos sólidos urbanos no Brasil: desafios tecnológicos, políticos e econômicos. Instituto de Pesquisa Econômica Aplicada (IPEA). Retrieved by <https://www.ipea.gov.br/cts/pt/central-de-conteudo/artigos/artigos/217-residuos-solidos-urbanos-no-brasil-desafios-tecnologicos-politicos-e-economicos>
- Bochkovskiy, A., Wang, C., & Liao, H. M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934. <https://doi.org/10.48550/arXiv.2004.10934>

- Filipakis, G., & Fagundes, F. (2021). Uma proposta de utilização de Redes Neurais na criação de sequências didáticas baseadas no padrão SCORM. In XXIII Congresso de Computação e Tecnologias da Informação – Encoinfo. Palmas. Retrieved from https://www.researchgate.net/publication/356223762_Uma_proposta_de_utilizacao_de_Redets_Neurais_na_criacao_de_sequencias_didaticas_baseadas_no_padrao_SCORM
- Geraeds, M., Van Emmerik, T., de Vries, R., & bin Ab Razak, M. S. (2019). Riverine Plastic Litter Monitoring Using Unmanned Aerial Vehicles (UAVs). *Remote Sensing*, 11, 2045. <https://doi.org/10.3390/rs11172045>
- Glanville, K., & Chang, H.-C. (2015). Remote Sensing Analysis Techniques and Sensor Requirements to Support the Mapping of Illegal Domestic Waste Disposal Sites in Queensland, Australia. *Remote Sensing*, 7, 13053-13069. <https://doi.org/10.3390/rs71013053>
- Green, D. R., Hagon, J. J., Gómez, C., & Gregory, B. J. (2017). Using Low-Cost UAVs for Environmental Monitoring, Mapping, and Modelling: Examples From the Coastal Zone. In A. G. De Luca, D. A. Waite, & P. J. Howarth (Eds.), *Remote Sensing for Coastal Environments* (pp. 425-448). Elsevier. <https://doi.org/10.1016/B978-0-12-810473-6.00022-4>
- Li, H., & Zhang, L. (2020). A Bilevel Learning Model and Algorithm for Self-Organizing Feed-Forward Neural Networks for Pattern Classification. *IEEE Transactions on Neural Networks and Learning Systems*, 32, 4901-4915. <https://doi.org/10.1109/TNNLS.2020.3026114>
- Lin, T. Y., Maire, M., Belongie, S. J., Bourdev, L. D., Girshick, R. B., Hays, J., ... Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. *CoRR*, abs/1405.0312. Retrieved from <http://arxiv.org/abs/1405.0312>
- Manfreda, S.; McCabe, M.F.; Miller, P.E.; Lucas, R.; Pajuelo Madrigal, V.; Mallinis, G.; Ben Dor, E.; Helman, D.; Estes, L.; Ciraolo, G.; et al. On the Use of Unmanned Aerial Systems for Environmental Monitoring. *Remote Sens.* 2018, 10, 641. <https://doi.org/10.3390/rs10040641>
- Mittal, G., Yagnik, K. B., Garg, M., & Krishnan, N. C. (2016). Spotgarbage: smartphone app to detect garbage using deep learning. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 940–945). <https://doi.org/10.1145/2971648.2971731>
- Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, Faster, Stronger. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 6517-6525). Honolulu, HI, USA. <https://doi.org/10.1109/CVPR.2017.690>
- Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*. <https://proceedings.neurips.cc/paper/2015/hash/14bfa6bb14875e45bba028a21ed38046-Abstract.html>
- Verma, V., Gupta, D., Gupta, S., Uppal, M., Anand, D., Ortega-Mansilla, A., ... Almotiri, J. (2022). A Deep Learning-Based Intelligent Garbage Detection System Using an Unmanned Aerial Vehicle. *Symmetry*, 14, 960. <https://doi.org/10.3390/sym14050960>
- Xu, D., & Wu, Y. (2020). Improved YOLO-V3 with DenseNet for Multi-Scale Remote Sensing Target Detection. *Sensors (Basel, Switzerland)*, 20. <https://doi.org/10.3390/s20154276>
- Youme, O., Bayet, T., Dembele, J. M., & Cambier, C. (2021). Deep Learning and Remote Sensing: Detection of Dumping Waste Using UAV. *Procedia Computer Science*, 185. <https://doi.org/10.1016/j.procs.2021.05.037>