

Machine learning e processamento digital de imagens: influência do pré-tratamento de dados no treinamento de um modelo para detecção de gates

Machine learning and digital image processing: influence of data preprocessing on training a model for gate detection.

Ryan Felipe Ferreira Ribeiro[†]

Thiago Batista Araújo[‡]

Resumo Este artigo destaca a influência do pré-processamento de dados no treinamento de uma inteligência artificial (IA) para detecção de gates. O estudo teve como objetivo comparar a acurácia e o tempo de treinamento entre imagens pré-processadas com filtro de cor laranja e imagens não pré-processadas.

A metodologia utilizada consistiu na criação de um *dataset* reduzido com 302 imagens, sendo metade pré-processada com o filtro de cor e a outra metade sem pré-processamento. Uma IA foi treinada usando Redes Neurais Convolucionais (CNN), utilizando 80% do *dataset* para treinamento e 20% para validação. A acurácia e a quantidade de ciclos necessários para o treinamento foram comparadas para analisar a influência do filtro de cor na detecção das gates.

Os resultados principais mostraram que as imagens pré-processadas com o filtro de cor laranja apresentaram uma maior acurácia na detecção das gates, além de uma redução na quantidade de ciclos necessários para o treinamento. A acurácia aumentou aproximadamente 16,1 pontos percentuais para o modelo treinado.

Como conclusão, observa-se que a utilização do filtro de cor é uma técnica válida para aumentar a acurácia na detecção de gates, desde que seja aplicada corretamente, evitando o *overfitting* do modelo. Esses resultados sugerem que o processo de treinamento de uma IA pode ser menos computacionalmente custoso se o pré-processamento digital da imagem for realizado adequadamente.

Palavras-chave: Visão Computacional; Inteligência artificial; Embarcados; Machine Learning; Processamento digital de imagens; Drones

Abstract This article highlights the influence of data preprocessing on training an artificial intelligence (AI) for gate detection. The study aimed to compare the accuracy and training time between preprocessed images with an orange color filter and non-preprocessed images.

The methodology involved creating a reduced dataset with 302 images, half of which were preprocessed with the color filter, while the other half remained non-preprocessed.

[†]Itajubá, Brasil: ferreira4191@gmail.com

[‡]Itajubá, Brasil: thiagobatistaaraujo06@gmail.com

A CNN (Convolutional Neural Network) was trained using 80% of the dataset for training and 20% for validation. The accuracy and the number of training cycles were compared to analyze the influence of the color filter on gate detection.

The main results showed that the preprocessed images with the orange color filter exhibited higher accuracy in gate detection, along with a reduction in the number of training cycles required. The accuracy increased by approximately 16,1 percentage points for the trained model.

In conclusion, it is observed that the use of the color filter is a valid technique to increase accuracy in gate detection, provided it is applied correctly to avoid model overfitting. These results suggest that the training process of an AI can be less computationally costly if digital image preprocessing is performed adequately.

Keywords: Computer Vision; Artificial Intelligence; Embedded Systems; Machine Learning; Digital Image Processing; Drones.

1. Introdução

O mercado e as aplicações que envolvem inteligência artificial e machine learning, segundo a International Data Corporation (IDC), estão em um processo de expansão. De acordo com a IDC Worldwide Semiannual Artificial Intelligence Tracker, espera-se um crescimento anual composto (CAGR) de 18,6% entre 2022 e 2026 para áreas que abrangem software, hardware e serviços de IA. Essa projeção indica que o valor estimado do mercado de IA poderá alcançar US\$ 900 bilhões até 2026 (1).

Nesse contexto em crescimento, a utilização da inteligência artificial em dispositivos como drones tem impulsionado o desenvolvimento de cenários de visão computacional. Essa abordagem envolve treinar máquinas para reconhecer objetos de maneira análoga à percepção visual humana. A aplicação da visão computacional em drones autônomos surge da necessidade desses dispositivos reconhecerem objetos em seu contexto, como gates, a fim de realizar movimentações precisas e evitar acidentes.

No entanto, o treinamento de uma IA pode ser custoso e demorado em termos de processamento. Portanto, explorar estratégias complementares, como o uso do processamento digital de imagem, pode ser uma abordagem válida para aprimorar a detecção de objetos. O processamento digital de imagem se destaca como uma ferramenta auxiliar no reconhecimento de objetos, permitindo melhorar a capacidade de detecção e noção de realidade de dispositivos e sistemas.

O presente artigo tem como objetivo analisar a influência da aplicação do processamento digital de imagem, particularmente do filtro de cor, em um drone equipado com uma IA previamente treinada para reconhecer gates. A hipótese subjacente a este estudo é que o uso de um simples filtro de cor pode ter um impacto significativo no aprimoramento do treinamento de um modelo, além do reconhecimento de gates pela IA. Acredita-se que ao aplicar um filtro

de cor específico, correspondente à coloração das gates, a IA será capaz de melhorar sua capacidade de identificar e localizar com precisão esses objetos.

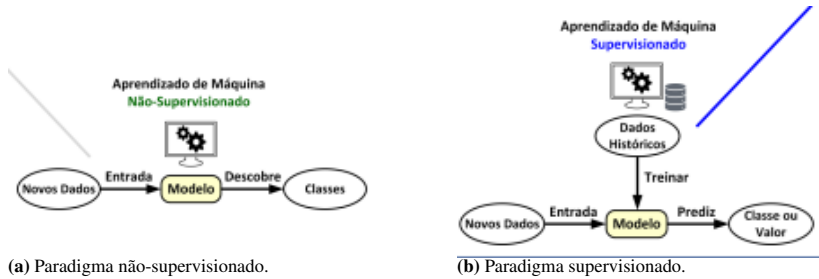
Para realizar este estudo, foi adotada uma metodologia que envolveu a criação de um *dataset* contendo imagens de gates em diferentes cenários. A partir desse conjunto de dados, foram selecionadas 302 imagens para o treinamento do modelo. Metade dessas imagens correspondeu às imagens originais, enquanto a outra metade foi processada aplicando-se o filtro de cor laranja, característico da coloração das gates. O filtro de cor foi implementado utilizando a biblioteca OpenCV em Python. A IA foi treinada com base nessas imagens selecionadas e processadas utilizando a plataforma Edge Impulse. Essa abordagem permite analisar de forma comparativa os resultados obtidos com e sem a aplicação do filtro de cor no reconhecimento das gates.

1.1 Machine Learning

O aprendizado de máquina (Machine Learning) é o principal campo de pesquisa da inteligência artificial, que foca no estudo de como as máquinas podem adquirir conhecimento e realizar comportamentos semelhantes aos seres humanos. Para que um problema seja computacionalmente tratável, os objetos reais devem ser representados por seus atributos em forma de dados. Esses atributos possuem tipos diferentes:

- Nominais: nomes (rótulos) que indicam categoria. Ex: “cor do cabelo” pode ser preto, castanho, etc.
- Numéricos: valores inteiros ou reais. Ex: peso e temperatura.

Como o objetivo é alcançar algo semelhante ao aprendizado humano, podemos observar como nós mesmos resolvemos problemas. Em uma situação hipotética, relacionando um ferro de passar roupa em alta temperatura e uma criança, possivelmente se a criança não for avisada ou o ferro não se encontrar longe de seu alcance, teremos como resultado uma queimadura. Isso ocorre pois a criança ainda não teve a experiência ou informação de que o ferro poderia queimar sua mão. Essa situação vale para pessoas de qualquer idade que quando se deparam com um problema inédito tentarão solucioná-lo baseado em algo que já ouviu ou lhe ocorreu. No aprendizado de máquina a situação é semelhante, no qual podemos ter dois tipos de paradigmas: o aprendizado supervisionado e não-supervisionado. No supervisionado, o modelo recebe dados históricos rotulados para aprender e inferir novos, enquanto no não supervisionado a tarefa do modelo é justamente rotular de acordo com algum critério de similaridade.

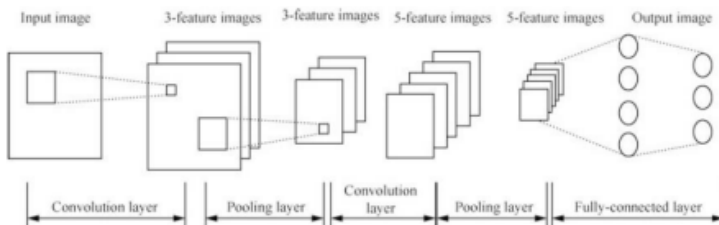
**Figura 1**

Comparação de paradigmas (Adaptado de Huawei Technologies Co. Artificial Intelligence Technology - Official Textbooks for Huawei ICT Academy, 3ª edição, Morgan Kaufmann, 2023).

Dentro dos paradigmas pode-se existir problemas que envolvem classificação de objetos, ou seja, onde o objetivo é rotular os novos dados de entrada, e problemas que envolvem regressão onde o objetivo é encontrar um valor ideal para uma equação, por exemplo, para encontrar as constantes em uma malha de controle proporcional integral derivativo (PID).

1.2 Deep Learning

Em aprendizado de máquina existem redes de treinamento que possuem várias camadas de cálculo e processamento. Esse conceito é utilizado onde é necessário extrair até o mínimo detalhe de um objeto para realizar a classificação, sendo necessário a presença de vários estágios de profundidade. No caso específico desta pesquisa, trabalhamos com Redes Neurais Convolucionais (CNN), que são voltadas para o processamento de imagens e visão computacional. Essa rede possui três tipos de camadas, sendo elas: convolucionais, *pooling* e densas.

**Figura 2**

Estrutura rede convolucional (Adaptado de Huawei Technologies Co. Artificial Intelligence Technology - Official Textbooks for Huawei ICT Academy, 3ª edição, Morgan Kaufmann, 2023).

Chama-se de convolucional a camada onde os neurônios dissecam características da imagem por partes, aprendendo padrões locais. Assim, em cada

etapa é aplicada a operação de covolução na imagem, utilizando de uma matriz filtro para extrair o que denominamos como *features*.

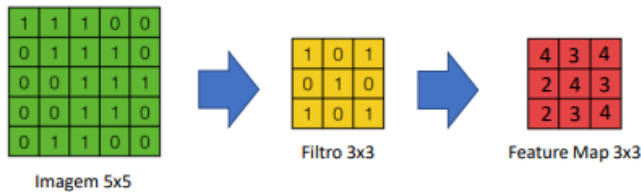


Figura 3

Extração de *features* camadas convolucionais (Adaptado de Huawei Technologies Co. Artificial Intelligence Technology - Official Textbooks for Huawei ICT Academy, 3ª edição, Morgan Kaufmann, 2023).

Após a extração das *features*, as mesmas passam pela camada de *pooling*, que reduz as dimensões, com intuito de diminuir o gasto computacional. Por fim, após várias camadas de convolução e *pooling*, as camadas densas servem como saída da CNN - como um classificador da última *feature* extraída.

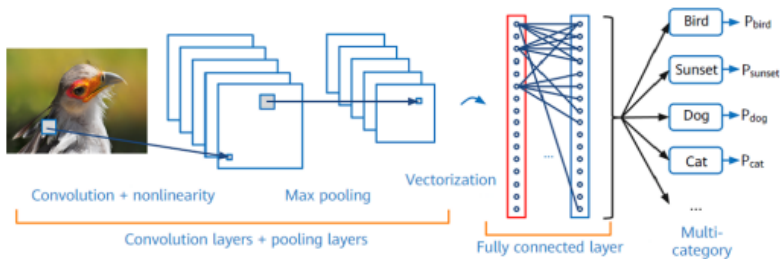


Figura 4

Rede convolucional (Adaptado de Huawei Technologies Co. Artificial Intelligence Technology - Official Textbooks for Huawei ICT Academy, 3ª edição, Morgan Kaufmann, 2023).

1.3 Visão Computacional

A visão computacional é uma área interdisciplinar que busca capacitar os computadores a compreender e interpretar informações visuais, assim como os seres humanos o fazem. Utilizando conhecimentos da física, óptica, matemática, estatística e inteligência artificial, a visão computacional desenvolve algoritmos e técnicas para analisar e extrair informações valiosas de imagens

e vídeos. É uma área que combina teoria e prática, buscando replicar a capacidade humana de interpretação visual e permitindo que as máquinas entendam e processem o mundo visual ao seu redor (3).

Ao treinar uma IA e embarcá-la em um dispositivo como drone, para que o mesmo passe a conseguir detectar objetos, estamos utilizando de conceitos da visão computacional, tendo em vista que, por meio de uma câmera, um algoritmo foi treinado para ser capaz de definir, com uma certa precisão, se o que está sendo captado nas imagens é ou não o alvo da busca. O processamento digital de imagem, também abordado na pesquisa, é uma ferramenta auxiliar à visão computacional.

2. Desenvolvimento

2.1 Metodologia e Procedimento

Nessa pesquisa optamos por realizar testes simplórios para a extração de resultados que não visassem a precisão, mas sim a comparação dos dois métodos, por isso trabalhamos com um banco de dados reduzido. Quando se trata de realizar um projeto de aprendizado de máquina, etapas devem ser seguidas à risca para garantir uma boa acurácia no resultado. Esse processo parte da coleta dos dados, ao tratamento desses dados, extração de características, criação do modelo, teste, validação e por fim a implantação. Existem várias plataformas que abstraem esse trabalho de treinamento em uma interface intuitiva, e nesse caso foi utilizada a plataforma Edge Impulse, que prove um espaço de trabalho gratuito (com certas limitações de uso em larga escala).

2.1.1 Coleta de Dados

Considerada uma das etapas mais importantes do treinamento, como diz o artigo publicado na MIT Sloan Review, a competição na área de aprendizado de máquina está realmente relacionada a uma corrida pelos dados, já que para o restante do processo temos ferramentas que abstraem o seu desenvolvimento (7). Como o objetivo é extrair dados do meio físico, a melhor escolha é o uso de sensores que podem variar desde sensores de distância, como *Light Detection and Ranging* (LIDAR) e ultrassônicos, até a própria câmera. No caso da pesquisa, utilizamos a câmera OAKD-Lite para tirar fotos das gates, visto que possui ampla compatibilidade com visão computacional e IA, na intenção de que em um trabalho futuro possamos fazer a implementação na própria câmera.

Na intenção de driblar o problema comum em modelos que não apresentam boas capacidades de generalização, ou seja, apresentam eficácia somente

para uma situação que estão treinados, as imagens foram capturadas em diferentes locais, posições e iluminações.



(a) Imagem capturada em ambiente de iluminação artificial e claro.



(b) Imagem capturada em ambiente de iluminação natural e com perspectiva de ângulo alto.



(c) Imagem capturada em ambiente de iluminação mista e com perspectiva de ângulo baixo.

Figura 5

Exemplos de imagens capturadas.

Esses dados coletados serão divididos em subseções de treinamento e validação, na qual um é dedicado ao aprendizado do modelo e o outro para testar a eficácia com dados desconhecidos. Uma abordagem comum, e que foi utilizada pelo próprio Edge Impulse é 80/20, na qual 80% dos dados são dedicados ao aprendizado e 20% à validação.

2.1.2 Tratamento de Dados

Além disso, é necessário que haja a manipulação desses dados visando o objetivo final. No caso desta pesquisa, como queremos identificar gates, é necessário rotular os dados de treinamento que servirão de base histórica para o modelo.

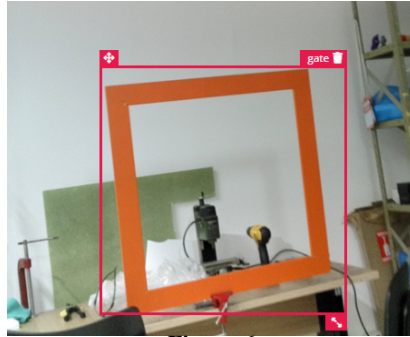


Figura 6
Processo de rotulagem.

Após isso, redimensionamos as imagens para um valor padrão, que no caso foi escolhido 96 pixels x 96 pixels. No entanto, como desejamos analisar também a vertente com imagens filtradas, foi utilizado um algoritmo em python que utiliza da biblioteca de processamento de imagens OpenCV para filtrar a cor laranja da imagem, tendo como objetivo uma figura preta e branca onde o branco representaria o alvo e o preto o plano de fundo. Vale-se ressaltar que não passamos o filtro perfeito para todas as imagens, já que o objetivo é justamente a IA auxiliar em esquivar dos ruídos presentes.

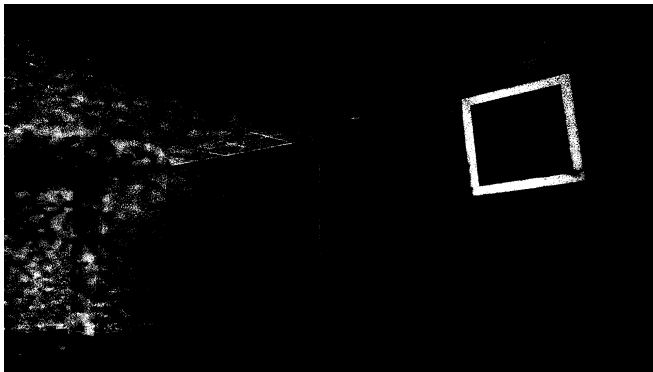


Figura 7
Imagem pré-processada.

Na Figura 7 pode-se observar pelas manchas brancas, que existe uma rede lateral que possui uma coloração laranja semelhante à cor da gate, e é justamente nesse problema que o modelo pretende atuar, deixando a detecção de cor por conta da visão computacional e agindo no padrão de formato e ignorando ruídos.

2.1.3 Extração de Características

Nessa etapa, a imagem é dissecada no que denominamos “*features*”, que são conjuntos de pequenos detalhes presentes no frame. A partir disso, as imagens podem ser reunidas em grupos que apresentam características mais semelhantes.

2.1.4 Treinamento do Modelo

Na seção de treinamento foram definidos os parâmetros como número de ciclos de treinamento, a taxa de aprendizado da rede (*learning rate*), se usaremos *data augmentation* (técnica que consiste em aumentar a amostra, aplicando inversões, filtros, rotações, para, assim, adquirir um maior poder de generalização). Para os dois experimentos foram definidos 50 ciclos, um *learning rate* de 0,001 e uso do *data augmentation*. E como esperado, realizamos o treinamento das imagens através das camadas definidas.

A estrutura de rede consiste de 27.648 *features* na entrada, o modelo *Faster Objects, More Objects* (FOMO) usado para distinguir objetos de interesse e plano de fundo usando de pouco processamento para isso, e na saída temos um *layer* para classificar se temos ou não gate.

2.1.5 Teste e Validação

É o momento de verificar o desempenho do modelo para dados desconhecidos que separamos anteriormente. A partir do modelo treinado, é solicitado que o mesmo faça a predição de cada uma das imagens de teste, verificando se encontra ou não o alvo de maneira correta.

2.2 Resultados e Discussões

2.2.1 Features

No caso das imagens sem pré-processamento, foi obtido o mapeamento de características semelhantes (capturadas no mesmo ambiente, iluminação e posição), dividida por lote. Enquanto isso, para o modelo pré-processado, foi obtido um mapeamento bem mais esparsso, como esperado, já que retirando o fundo analisou-se também uma redução de características usadas para separação em grupos. Sendo assim, não interessa o ambiente e iluminação, uma vez que o modelo só reconhece o “mundo filtrado” que lhe interessa e alguns ruídos.

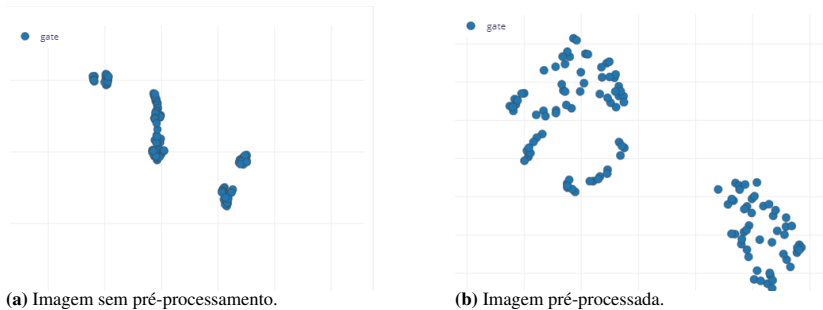


Figura 8
Mapeamento de features.

2.2.2 Treinamento

Pode-se notar a partir das imagens, que o modelo com imagens pré-processadas atingiu uma escala de acurácia maior. Isso revela que podemos reduzir o número de ciclos de treinamento para atingir um objetivo, o que para poucos dados pode parecer irrelevante, mas quando se trata de um *dataset* significativamente grande, pode economizar em gasto computacional. Além disso, esse fator se mostra essencial quando se busca um *deploy* em um dispositivo embarcado com sistema limitado.

Contudo, é necessário se atentar ao fator de *overfitting* (condição em que o modelo aprende demais com aquele conjunto de dados e perde capacidade de generalizar), que pode ocorrer mais rápido nas imagens filtradas, já que reduzimos o campo de características a serem aprendidas.

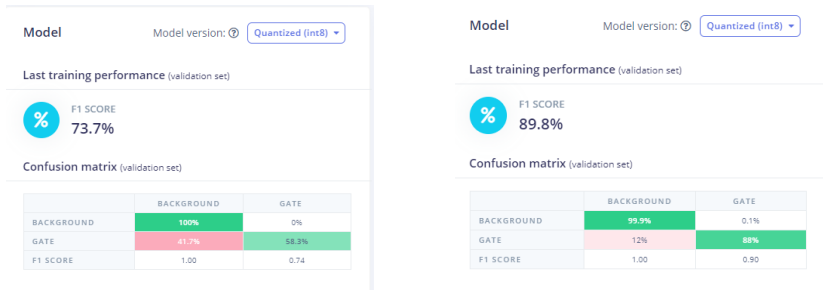


Figura 9
Treinamento do modelo.

2.2.3 Validação

Em quesito de validação, a diferença se mostra ainda maior com a imagem pré-processada obtendo êxito nas inferências realizadas.

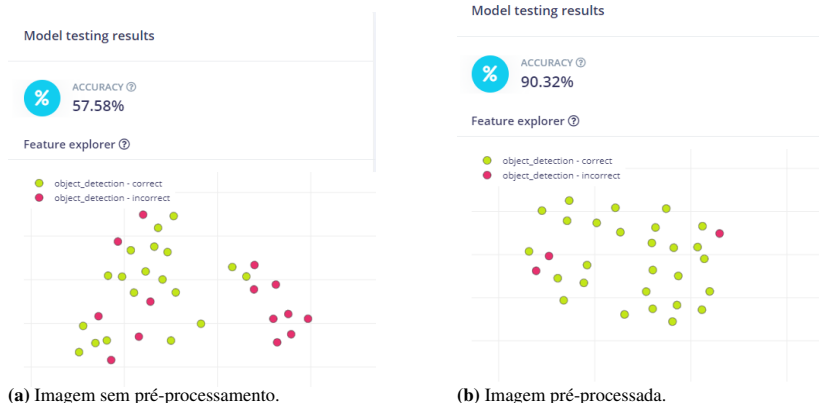


Figura 10
Validação do modelo.

3. Conclusão

A junção do treinamento de uma inteligência artificial com um pré-processamento de uma imagem é uma solução válida, na visão computacional, para se detectar objetos específicos.

Com base nos resultados obtidos nesta pesquisa, pode-se concluir que o uso do filtro de cor tem um impacto significativo na melhoria da precisão da detecção e na redução do número de ciclos de treinamento da IA. A aplicação desse filtro no processamento digital de imagens, especificamente na detecção de uma gate, mostrou-se eficaz ao melhorar significativamente a acurácia do modelo.

Essa descoberta tem implicações importantes no campo da visão computacional pois a relevância desse estudo reside no fato de que, em cenários específicos, como na detecção de uma gate, o processo de treinamento da IA se torna menos exigente em termos computacionais, ao mesmo tempo em que melhora a precisão do modelo. Apesar disso, deve-se levar em conta fatores como *overfitting*, ou seja, a utilização do pré-processamento deve ser cauteloso e testado.

Vale ressaltar que a quantidade de imagens testadas foi limitada, o que pode ter afetado a robustez do modelo em si. Para futuras direções de pesquisa, pretendemos aprimorar o modelo por meio do treinamento com um

dataset maior e mais diversificado, visando melhorar ainda mais a acurácia e generalização do sistema. Além disso, pretendemos integrar esse modelo a uma plataforma funcional, como a OAKD-lite, para expandir seu uso e avaliar seu desempenho em uma aplicação real.

Referências

- [1] IDC. *IDC Worldwide Semiannual Artificial Intelligence Tracker*. Disponível em: <https://www.idc.com/getdoc.jsp?containerId=prEUR249536522>. Acesso em: 9 de junho de 2023.
- [2] Huawei Technologies Co. *Artificial Intelligence Technology - Official Textbooks for Huawei ICT Academy*. 3ª edição. Morgan Kaufmann, 2023.
- [3] Szeliski, R. (2010). *Computer Vision: Algorithms and Applications*. Springer Science & Business Media.
- [4] Edge Impulse. *Edge Impulse*. Disponível em: <https://www.edgeimpulse.com>. Acesso em: 12 de junho de 2023.
- [5] Müller, A.C. and Guido, S. *Introduction to Machine Learning with Python: a guide for data scientists*. O'Reilly Media, Inc., 2016.
- [6] PyImageSearch. *PyImageSearch*. Disponível em: <https://pyimagesearch.com>. Acesso em: 12 de junho de 2023.
- [7] Beck, M., Libert, B. (n.d.). The Machine Learning Race Is Really a Data Race. *MIT Sloan Review*. Retrieved from <https://sloanreview.mit.edu/article/the-machine-learning-race-is-really-a-data-race/>
- [8] Li, S., Horst, E., Dürnay, P., De Wagter, C., Croon, G. (2019, May). Visual Model-predictive Localization for Computationally Efficient Autonomous Racing of a 72-gram Drone. Retrieved from https://www.researchgate.net/publication/333394180_Visual_Model-predictive_Localization_for_Computationally_Efficient_Autonomous_Racing_of_a_72-gram_Drone