



UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA

Autorizada pelo Decreto Federal nº 77.496 de 27/04/76
Recredenciamento pelo Decreto nº 17.228 de 25/11/2016



PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
COORDENAÇÃO DE INICIAÇÃO CIENTÍFICA

XXVII SEMINÁRIO DE INICIAÇÃO CIENTÍFICA DA UEFS SEMANA NACIONAL DE CIÊNCIA E TECNOLOGIA - 2023

DESENVOLVIMENTO DE UM SISTEMA DE VISÃO PARA DETECÇÃO DE OBJETOS EMBARCADO EM UM BRAÇO ROBÓTICO.

Igor Figueredo Soares¹; Anfranserai Morais Dias².

1. Bolsista PROBIC/UEFS, Graduando em Engenharia da Computação, UEFS, e-mail: ifs54@hotmail.com

2. Orientador, Departamento de Tecnologia, UEFS, e-mail: anfranserai@ecompu.uefs.br

PALAVRAS-CHAVE: sistema de visão; detecção de objeto; TensorFlow.

INTRODUÇÃO

Sistemas robóticos autônomos necessitam compreender o ambiente à sua volta para tomada de decisões. Dentro do contexto de braços robóticos, o reconhecimento de objetos é uma atividade fundamental para as operações de agarramento e navegação (Kumar, 2014). A visão computacional proporciona ferramentas que processam e extraem informações de imagens. Uma dessas ferramentas tem como propósito detectar e classificar objetos presentes em uma imagem, de maneira independente, por meio de modelos de aprendizado profundo (*deep learning*) (Brownlee, 2019). O presente trabalho visa desenvolver um sistema de visão para a detecção de objetos a ser embarcado em um braço robótico. O sistema faz uso do módulo ESP32-CAM para capturar as imagens e um sensor de distância para identificar a distância até os objetos. Os dados coletados pelo módulo são transmitidos para um computador via rede sem fio. A partir disso, são analisados para a identificação da presença do objeto de interesse e suas dimensões (largura e altura).

MATERIAL E MÉTODOS

O módulo de aquisição foi embarcado no ESP32-CAM, que funciona como um servidor fornecendo as imagens e a distância do objeto para o módulo de inferência e mensuração. Para aquisição das imagens utilizou-se um sensor OV2640 de 2MP. Enquanto que para obter a distância foi utilizado o sensor VL53L0X. Devido a limitações de processamento optou-se por utilizar a resolução 640x480 pixels. O módulo de inferência e mensuração é executado em um computador convencional. A principal atribuição do módulo é efetuar a identificação do objeto na imagem recebida por meio de um Modelo de Detecção de Objetos. Caso o objeto seja identificado, as suas dimensões (em pixels) são submetidas a um Modelo de Mensuração de Objetos que determina o seu tamanho real.

Modelo de Detecção de Objetos

O modelo de detecção de objetos foi construído com base no modelo de arquitetura EfficientDet-Lite, utilizando a técnica de transferência de aprendizado. O desenvolvimento do modelo foi feito por meio da biblioteca TensorFlow e exportado

para o formato Tensorflow Lite. O EfficientDet-Lite é um modelo derivado do EfficientDet, treinado com o dataset COCO 2017, otimizado para o Tensorflow Lite e desenvolvido para dispositivos móveis (Kamal; Suciati; Hidayati, 2021). O EfficientDet é um modelo de detecção de objetos que utiliza o EfficientNet como rede principal combinada com a Bidirectional Feature Pyramid Network (BiFPN). A BiFPN atua como uma rede de características, e aplica a fusão de características bidirecionais de cima para baixo e de baixo para cima. O resultado dessas características são inseridos na rede de classes e caixas para produzir a previsão das classes e as caixas delimitadoras dos objetos na imagem (Tan; Pang; Le, 2020).

A técnica de transferência de aprendizado permite o aproveitamento de pesos aprendidos por um modelo treinado com um grande conjunto de dados para uma determinada tarefa, na construção de um outro modelo que execute uma tarefa similar. Essa abordagem possibilita treinar modelos com conjunto de dados (dataset) menores e com menos recursos computacionais.

Modelo de Mensuração de Objetos

O modelo foi construído com base na relação entre o tamanho de um objeto no mundo real (em mm) e na imagem (em pixels), a uma distância conhecida. Deste modo, foram efetuadas capturas de imagens com o ESP32-CAM das peças de legos a uma distância conhecida em relação ao microcontrolador. Foi efetuado um conjunto de capturas para 20 distâncias distintas, iniciando em 50 mm e variando de 50 em 50 mm até atingir a distância de 1000 mm. A partir das imagens coletadas foram calculadas as médias das dimensões em pixels das peças para cada uma das distâncias. Utilizou-se uma regressão linear para obter equações que estimam a relação mm/pixel das dimensões em função da distância em que a peça se encontra do protótipo.

RESULTADOS E DISCUSSÃO

Para a construção do sistema, decidiu-se por utilizar um conjunto de peças de legos como objeto de interesse. A adoção de peças de legos como objeto de interesse se deve ao fato de proporcionarem diferentes arranjos por meio da combinação das peças, permitindo que tivéssemos um mesmo objeto com diferentes tamanhos. O dataset utilizado para treinar o modelo foi construído por meio da aquisição das imagens das peças de legos utilizando o módulo ESP32-CAM. O conjunto de dados conta com 1020 imagens, onde 90% foi destinada ao treinamento do modelo e 10% para validação. A figura 1 mostra algumas imagens do conjunto de dados.



Figura 1: Exemplo de imagens presentes no conjunto de dados.

A avaliação do modelo foi feita seguindo as métricas definidas pelo método Common Objects in Context (COCO) e com um conjunto de teste de 236 imagens (COCO, 2015).

Os componentes de aquisição do sistema foram acoplados a uma placa de circuito, formando um conjunto único para aquisição dos dados como mostra a figura 2.

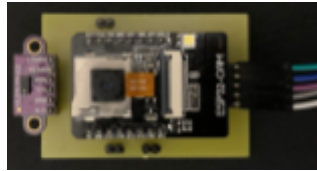


Figura 2: Sistema de aquisição.

Detecção de Objetos

Para o treinamento do modelo de detecção de objetos, a rede neural percorre o conjunto de dados 20 vezes e processa quatro amostras antes dos parâmetros internos da rede serem atualizados. Após o treinamento, o modelo foi exportado para o formato do Tensorflow Lite. A análise do desempenho do modelo utiliza o conjunto de dados de validação e um conjunto de dados de teste. O conjunto de dados de teste é composto por um grupo de imagens que simulam diferentes situações em que o sistema poderia ser submetido, como por exemplo, ambientes com pouca ou muita iluminação, ambientes internos e externos. A tabela 1 mostra todos os valores de desempenho do modelo de acordo com as métricas COCO.

Tabela 1 - Desempenho do modelo.

Métricas Avaliadas											
AP	AP ^{IoU=0,50}	AP ^{IoU=0,75}	AP ^{small}	AP ^{medium}	AP ^{large}	AR ^{max=1}	AR ^{max=10}	AR ^{max=100}	AR ^{small}	AR ^{medium}	AR ^{large}
Conjunto de Dados de Validação											
81,92	95,69	92,89	52,35	79,81	92,05	76,75	85,18	85,18	59,23	84,26	93,40
Conjunto de Dados de Teste											
61,86	87,41	70,62	34,05	68,28	77,61	64,33	69,05	69,54	46,23	73,88	84,06

O modelo apresentou um pior desempenho com o conjunto de dados de testes se comparado com os de validação. Possivelmente, isto se deve ao fato de que os dados de teste possuem imagens em ambiente não estruturado.

Mensuração de Objetos

A avaliação do sistema de mensuração foi feita com o sistema já acoplado ao modelo de detecção de objetos desenvolvido. Foram utilizados dois conjuntos de legos, um com altura de 11 mm e largura de 31,5 mm e outro com altura 21 mm e largura de 31,5 mm. Foram coletadas imagens dos conjuntos a distâncias conhecidas e calculada a raiz quadrada do erro médio (RMSE). Para o primeiro conjunto de legos o RMSE da altura foi de 2,73 mm e da largura 2,22 mm. Já para o segundo conjunto, o RMSE da altura foi de 2,46 mm e da largura de 2,39 mm.

O segundo teste avalia o comportamento do modelo para objetos em perspectiva. Para isso, o modelo foi submetido a três conjuntos de legos de tamanhos distintos, dispostos de modo em que suas faces não estivessem em paralelo ao plano da imagem. O modelo apresenta comportamentos distintos para objetos em perspectiva, ora apresentando estimativas congruentes, ora apresentando estimativas errôneas. Deste modo, podemos inferir que o modelo não é o mais adequado para se trabalhar com objetos em perspectiva, mas sim, com objetos em que alguma de suas faces esteja paralela ao sistema de visão.

Sistema Completo

A análise de performance do protótipo foi feita utilizando dois blocos distintos: um com altura de 98,0 mm e largura de 31,5 mm; o outro com altura de 59,0 mm e largura de 31,5 mm. Os blocos foram posicionados a distâncias que variaram entre 40 e 540 mm. Não foi possível avaliar para distâncias menores porque o objeto não era enquadrado na imagem devido a proximidade com o sensor de captura.

A partir de 380 mm o RMSE da altura apresenta valores elevados (acima de 10 mm) em ambos os blocos. Isso é um reflexo das limitações do sensor de distância para obter distâncias até objetos de tamanho reduzido. Para distâncias maiores que 340 mm o sensor apresentou valores distorcidos da realidade, prejudicando o desempenho do sistema. Deste modo, o sistema se torna limitado na sua distância de atuação, garantindo o melhor desempenho até distâncias de 340 mm.

CONSIDERAÇÕES FINAIS (ou Conclusão)

O sistema de detecção de objetos construído é capaz de desempenhar as atribuições para o qual foi desenvolvido. Contudo, ainda apresenta limitações na mensuração do tamanho dos objetos a determinadas distâncias e na detecção de objetos em algumas condições, além de depender de um computador externo para efetuar as inferências.

Deste modo, em trabalhos futuros, o sistema pode ser aperfeiçoado com a construção de um novo modelo que possa ser embarcado no microcontrolador. Além disso, outros sensores de distância podem ser empregados com a finalidade de permitir que o sistema seja capaz de atuar em uma gama maior de situações.

Por fim, o plano de trabalho foi extremamente importante para o entendimento de como funciona todo o processo de pesquisa científica, além de proporcionar a aquisição de conhecimentos com aspectos práticos.

REFERÊNCIAS

- BROWNLEE, J. 2019. Deep learning for computer vision: image classification, object detection, and face recognition in python. Machine Learning Mastery.
- COCO. 2015. Detection evaluation metrics. Disponível em: <https://cocodataset.org/#detection-eval>. Acesso em: 17 ago. 2023.
- KAMAL, M.U.A.; SUCIATI, N.; HIDAYATI, S.C. 2021. Calculation of Handwriting Mathematics Expressions on Mobile Devices Using *Efficientdet-Lite0* and Reverse Polish Notation. The Journal on Machine Learning and Computational Intelligence (JMLCI), v. 1, n. 1. DOI: <https://doi.org/10.26740/volliss1y2021id4>.
- KUMAR, R. et al. 2014. Object detection and recognition for a pick and place robot. In: Asia-Pacific world congress on computer science and engineering. IEEE. p. 1-7. DOI: <https://doi.org/10.1109/APWCCSE.2014.7053853>.
- TAN, M.; PANG, R.; LE, Q.V. 2020. Efficientdet: Scalable and efficient object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. p. 10781-10790. DOI: <https://doi.org/10.48550/arXiv.1911.09070>.