

CONTAGEM DE PEDESTRES BASEADA NA DETECÇÃO DE CENTRÓIDES DE OBJETOS

Dielson Soares de Oliveira Junior¹ (EG), Egon Luiz Muller (PQ)¹

¹Universidade Federal de Itajubá - Campus Itajubá.

Palavras-chave: Inteligência Artificial, Machine Learning, TinyML, Vigilância Autônoma.

Introdução

Neste trabalho, objetiva-se detectar e contar humanos em uma imagem com o algoritmo de aprendizado de máquina FOMO (*Faster Objects, More Objects*) da plataforma Edge Impulse. Utilizou-se este algoritmo pela leveza teórica em memória e sua velocidade de execução, o que abre oportunidades para aplicações mais complexas em equipamentos mais simples. Esta leveza se deve ao algoritmo buscar os centróides dos objetos, ao invés de marcar uma área em que o objeto se encontra, como em outros algoritmos de detecção de objeto (MOREAU; KELCEY, 2022). Uma aplicação direta é a contagem de humanos em uma determinada imagem, que, por si só, contém diversas aplicações em segurança e controle de pessoal. A inteligência artificial foi treinada utilizando a própria plataforma Edge Impulse e, como base, foram utilizados *datasets* já existentes de pedestres. Os dados foram divididos entre treinamento, validação e teste, para que a inteligência artificial fosse avaliada de maneira apropriada.

Metodologia

Inicialmente, decidiu-se que a aquisição de dados seria feita a partir de coleções já existentes, testadas e replicadas em projetos diferentes. Para tal, escolheram-se os seguintes *datasets*: Multi-Person Tracking from Mobile Patterns, de ETH Zurich (ESS et al., 2008) e EuroCity Persons Dataset, de TU Delft (BRAUN et al., 2019). Estes dados foram então importados para a plataforma Edge Impulse e divididos aleatoriamente pela plataforma entre treinamento, validação e teste para trabalho posterior.

As imagens dos *datasets* foram então redimensionadas, cortadas e filtradas para se tornarem quadrados de resolução 320x320 em preto e branco, em função de diminuir o peso final da detecção.



Figura 1 – Imagem de exemplo do *dataset* filtrado
Com as imagens filtradas, a inteligência foi treinada utilizando o FOMO baseado em MobileNetV2 0.35. Para fins comparativos, outra inteligência foi treinada, baseada em MobileNetV2 SSD FPN-Lite, que é o competidor mais próximo do FOMO, com o mesmo *dataset*, porém, com as imagens na coloração original, que é a limitação deste algoritmo de aprendizado de máquina.

Resultados e discussão

Após o treinamento, a plataforma automaticamente executou testes para validação da inteligência. Abaixo, tabelou-se as saídas para cada gabarito:

Tabela 1 – Resultados dos testes com FOMO

GABARITO	RESULTADO	
	FUNDO	PESSOA
FUNDO	99,9 %	0,1 %
PESSOA	23,3 %	76,7 %

Com estes dados, entende-se que a inteligência é improvável de apontar falsos positivos, ou seja, detectar uma pessoa onde existe um fundo. Porém, 23,3% das vezes que há humanos em uma imagem, eles não são detectados. O que não significa que em uma imagem 1/4 (um quarto) das pessoas não são detectadas, mas sim que a inteligência falhou em contabilizar corretamente a quantidade de pessoas em apenas 23,3% das imagens. O ponto mais alto de utilização de memória volátil (RAM) foi 2,4 MB e o espaço de armazenamento ocupado pelo algoritmo de detecção é igual a 73,8 kB.

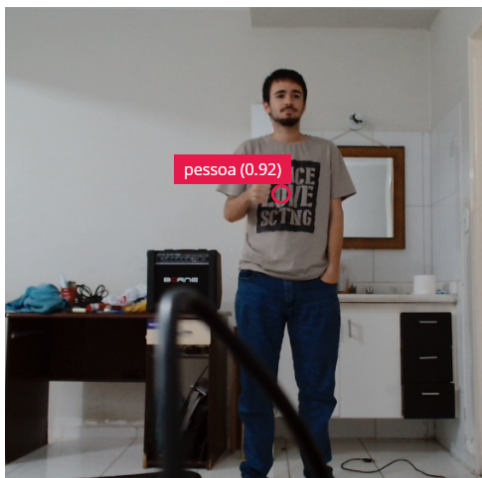


Figura 2 – O modelo detectando corretamente em um ambiente caótico desconhecido

Deste trabalho, fica claro que as soluções da Edge Impulse vão além das suas intenções originais, que são os processadores embarcados. Neste exemplo, de aplicação em vigilância, vê-se a possibilidade de utilizar as inteligências construídas através de seus algoritmos de aprendizado de máquina em projetos de larga escala, com processadores de ponta. Para trabalhos futuros, indica-se a expansão dos *datasets* e obtenção de dados próprios, com indivíduos inseridos no contexto brasileiro ao invés de europeu, como nos dados utilizados neste.

Agradecimento

Os principais agradecimentos se direcionam aos colegas e professores do PET-TEC, da Black Bee Drones e da UNIFEI que me proporcionaram diversas oportunidades. Também agradeço ao FNDE pelo apoio ao projeto.

Referências

BRAUN, Markus; KREBS, Sebastian; FLOHR, Fabian; GAVRILA, Dariu. EuroCity Persons: A Novel Benchmark for Person Detection in Traffic Scenes. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Anchorage, v. 41, n. 8, p. 1844-1861, ago. 2019.

ERIKSSON, Karl. Detecting Parasites on Bees with Mobile Object Detection. **Independent thesis**, Örebro, v. 1, n. 8, p. 47, out. 2021.

ESS, Andreas; LEIBE, Bastian; SCHINDLER, Konrad; GOOL, Luc van. A Mobile Vision System for Robust Multi-Person Tracking. **2008 IEEE Conference on Computer Vision and Pattern Recognition**, Anchorage, p. 1-8, jun. 2008.

MOREAU, Louis; KELCEY, Mat. Announcing FOMO (Faster Objects, More Objects). **Edge Impulse**, San Jose, mar. 2022. Disponível em: <<https://www.edgeimpulse.com/blog/announcing-fomo-faster-objects-more-objects>> Acesso em: 27/09/2022.

Adiante, os resultados dados pela detecção com o MobileNet:

Tabela 2 – Resultados dos testes com MobileNet

GABARITO	RESULTADO	
	FUNDO	PESSOA
FUNDO	100,0 %	0,0 %
PESSOA	28,7 %	71,3 %

Como se pode ver, os resultados da detecção com o MobileNet, um método mais robusto, não são muito diferentes dos resultados com o FOMO. Isso se deve, em grande parte, ao fato de que o FOMO é derivado do MobileNet.

A memória RAM máxima utilizada não é fornecida pelo Edge Impulse para o MobileNet, porém, sabe-se que os modelos treinados com esse algoritmo chegam a 30 MB de RAM (ERIKSSON, 2021), mais do que doze vezes maior do que o FOMO. O peso do modelo em armazenamento é de 11 MB, ou seja, quase cento e cinquenta vezes maior do que o FOMO.

Em vista destes resultados, é possível afirmar que o algoritmo de detecção pode ser utilizado em um computador pessoal de memória RAM maior ou igual a 2 GB para classificar dezenas de câmeras diferentes sem grandes esforços. Essas classificações podem ser utilizadas para gerar mapas de calor em ambientes que precisam de algum grau de controle de pessoas e uma grande área, como uma universidade, um espaço de festas, um parque de diversões ou até mesmo as vias públicas.

Conclusões