

USO DE UMA CNN PARA MAPEAMENTO DE BIOINDICADORES EM AMBIENTE DE PESQUISA LABORATORIAL

Wilgner Luis Pereira Barreto (IC)¹, Giovani Bernardes Vitor (PQ)², Fabrício Vieira De Andrade (PQ)³, Paulo José Lage Alvarenga (PQ)⁴

¹ UNIFEI, ² FAPEMIG

Resumo

O rastreamento de objetos é a base de muitos sistemas que fazem uso de Inteligência Artificial. Embora tenham sido feitos progressos consideráveis nesta área nos últimos anos, o rastreamento eficiente e preciso em vídeo continua sendo um desafio. Neste trabalho, é apresentado um rastreador que aproveita as informações de movimento do fluxo de vídeo compactado e um objeto semântico de uso geral com o detector atuando em quadros (frames). A abordagem proposta é testada em vídeos para medir sua eficiência autônoma. Resultados indicam potencial do método proposto em termos de precisão. Outras características desejáveis da proposta método são sua simplicidade e eficiência de implantação, que decorre do fato de reutilizar recursos e informações que podem já existir no sistema por motivos prévios.

Palavras-chave: Inteligência Artificial. Rede Neural Convolutacional. Sistemas. Distância Euclidiana, YOLO.

Introdução

Atualmente o reconhecimento e processamento de imagens é um tema de grande importância no terreno da Inteligência Artificial (IA), porém ainda encontra-se longe da perfeição. Cenários aparentemente simples, como detecção de objetos, reconhecimento facial, remoção de desfoque de movimento, etc. e cenários mais complexos, como artefatos de compressão, detecção de arranhões, ruído de capturas contínuas e detecção de alterações são aplicações de reconhecimento e processamento de imagem.

As imagens digitalizadas são frequentemente representadas como uma matriz bidimensional (2D) de valores de pixels. Cada valor de pixel que compõe o esquema de cores da imagem é frequentemente influenciado por uma matriz de fatores como a intensidade da luz. A cena visual é projetada contra uma superfície onde os receptores (naturais ou artificiais) produzem valores que dependem da intensidade da luz incidente (1).

O rastreamento é uma das tecnologias necessárias para o presente projeto que visa observar o comportamento de bioindicadores submersos (as artêmias salinas) que são utilizados em procedimentos de análise de águas para apontar o nível de poluição da mesma. O acompanhamento de vários objetos em vídeos requer a detecção de objetos em *frames* individuais e combinando a informação presente em cada um deles no final do processo. Muitas técnicas de visão computacional têm

sido utilizadas para construir Sistemas MOT (Multiple Object Tracking), e dia a dia a tecnologia cresce para fornecer rapidamente uma área de oportunidades chamada processamento de imagem (2).

Posteriormente, utiliza-se um software com instruções de mapeamento para armazenar e seguir estes bioindicadores alocados em uma placa de petri submersos por alguma substância de estudo, e a partir deste é gravado um vídeo de amostras de substâncias diferentes com objetivo de quantificar o tamanho da colônia viva ao fim do experimento. Ambas as abordagens são necessárias para quantificar e verificar a existência ou não de artêmias vivas em determinados experimentos.

Metodologia

A primeira questão da pesquisa é verificar a qualidade da Rede Neural chamada YOLO (*You Only Look Once*) como modelo para detecção geral dos objetos de interesse devido à natureza dos objetos classificados nesta pesquisa e ao fato de que este tipo de classificação ainda não existir, se YOLO é um potencial candidato para a detecção dos objetos pesquisados de forma a possibilitar um bom mapeamento.

A segunda questão da pesquisa é avaliar a precisão do modelo de detecção YOLO e o modelo de mapeamento sequencial de imagens usando as informações obtidas através do modelo treinado.

A seção a seguir apresenta conceitos, exemplos e aplicações do algoritmo implementados para a detecção e

também uma representação do aparato construído para capturar as imagens do objeto de estudo.

Protótipo de Captura de Imagens

Tendo em vista a natureza do tipo de ser vivo a ser observado foi necessário montar um observador que proporciona um espaço não influenciado por luzes externas ou perturbações por reflexão. Para isso, foi construído um objeto com luz de fundo estabilizada e em sua parte superior, uma câmera da marca Logitech, modelo C920S. O Objeto descrito acima pode ser visto na Figura 1.

Figura 1 – Aparato utilizado para efetuar a captura dos vídeos de estudo (produção dos autores).



Fonte: Autoria própria, 2022.

YOLO

YOLO é um algoritmo de detecção de objetos direcionado para processamento em tempo real. Ele difere de outros detectores de objetos usando uma única Rede Neural Convolutiva (CNN a sigla em inglês) para ambas as classificações e localização dos objetos (3).

O YOLO é implementado como uma rede neural de convolução, consiste em um total de vinte e quatro camadas convolucionais seguidas por duas camadas totalmente conectadas. As camadas são separadas por suas funcionalidades em relação à saída final da rede.

Treinamento da Rede YOLO

Modelos detectores de objetos são frequentemente empregados para analisar imagens do mundo real. Essas imagens são altamente complexas, sujeitas a várias condições de luz, ruídos da câmera, oclusões e outras condições ambientais (3). Um modelo de detector de

objetos precisa ser treinado com várias imagens capturadas sob uma ampla faixa percebida de condições ambientais para reconhecimento de objetos.

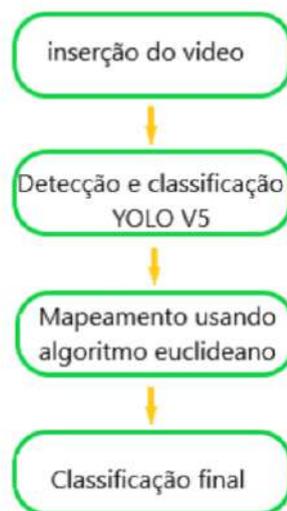
Um treinamento típico de rede de detector de objetos envolve quatro etapas principais. Selecionar uma arquitetura de rede, compilar dados de treinamento, treinar uma rede e testar o desempenho da rede treinada. Abaixo, estão individualmente descritas as etapas no contexto da rede YOLO v5 para reconhecimento de artemias.

Sistema de mapeamento temporal de objetos em vídeo

Ao se observar os resultados apresentados na seção anterior, nota-se que, quando a rede faz uma inferência de resultados, valores de confiabilidade são disponibilizados, ou seja para alcançar maior precisão em todas as detecções seria necessário uma eliminação de muitos 'candidatos' (5). Essa eliminação acaba diminuindo a eficiência da detecção para vários objetos de um mesmo rótulo, este que é o principal objetivo do presente projeto, portanto é necessário produzir uma forma de contornar este tipo de situação.

A Figura 2 demonstra um Diagrama simplificado de uma solução para diminuir a limitação de detecção utilizando alguns algoritmos baseados em coordenadas Euclidianas.

Figura 2 – Diagrama Básico do funcionamento da solução proposta.



Fonte: Autoria própria, 2022.

Considerando um vídeo como uma coleção de imagens subsequentes em um determinado espaço de tempo, podemos utilizar a causalidade das próximas imagens para definir uma continuação ou não de uma detecção de

uma artêmia, este processo nos algoritmos de visão são conhecidos como *Tracking* (2), ou acompanhamento.

Segundo Danielsson (6) um mapa indica, para cada pixel nos objetos (ou no fundo) da imagem originalmente binária, a distância mais curta até o pixel mais próximo no fundo (ou os objetos) pode ser descrito como uma distância euclidiana em pixels, permitindo um bom mapeamento de uma situação ao se somar as 3 imagens subsequentes, permitindo um mapeamento da mesma. A descrição pode ser vista a seguir.

$$D = \min(\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)}) < D_{MIN} \quad (a)$$

Onde: x_i e y_i , posição em pixels de uma determinada detecção a ser comparada com alguma outra detecção subsequente. D_{MIN} , valor máximo de admissão de subsequência.

Utilizando estes critérios, foi possível formalizar um critério geral para mapear as posições em um vídeo, e a partir destes poder aplicar o algoritmo ao ambiente proposto anteriormente.

Resultados e discussão

A pesquisa apresentou um estudo sobre a detecção e tracking de objetos por meio de uma rede neural convolucional. Em vários vídeos, o sistema proposto foi testado. O experimento é dividido em duas seções, a identificação e rastreamento de objetos. O design do projeto é baseado em Python e avaliados utilizando métricas padronizadas de desempenho de redes neurais.

Utilizando redes YOLO, o treinamento possui três métricas principais de desempenho que permitem ao algoritmo fazer novas detecções, são: **Precisão**, **Recall** e **mAP** (7).

A precisão é uma métrica de acurácia sobre os objetos preditos pela rede em relação aos reais. Ela visa apenas medir o quão bom o modelo estimado é em encontrar os verdadeiros objetos.

$$Precisão - P = \frac{TP}{TP + FP} \quad (b)$$

Recall é utilizado para medir o quão bom os objetos são encontrados pela rede. Significa dizer que o recall calcula a proporção de verdadeiros positivos (TP) que foram corretamente detectados.

$$Recall - R = \frac{TP}{TP + FN} \quad (c)$$

E por último, o mAP, é uma métrica que combina as métricas de precisão e de recall para definir a acurácia do objeto encontrado, este cálculo ocorre com a média de

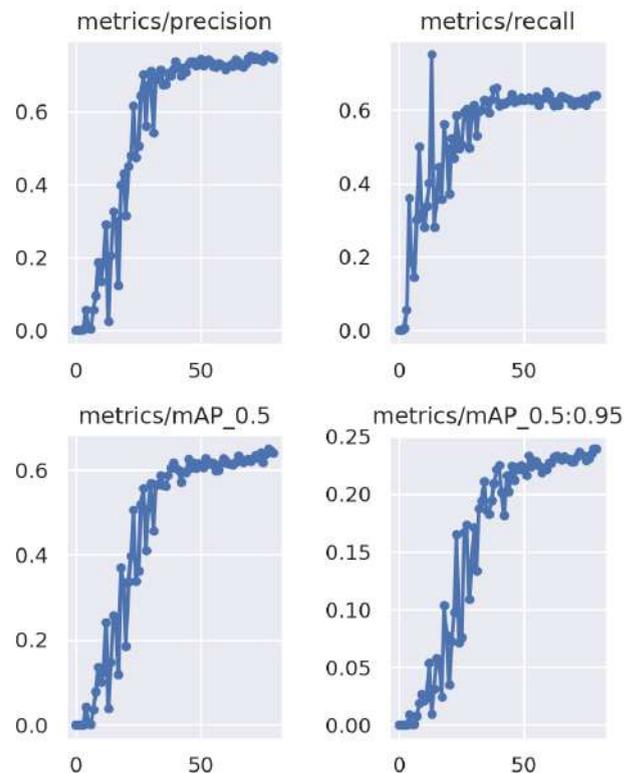
precisão e o valor de recall como sendo uma exigência mínima de confiabilidade. Este cálculo é feito em duas formas, com confiabilidade acima de 0 e outro entre 0,5 e 0,95.

$$AP = \frac{\sum_{i=1}^R P_i}{R}$$

$$mAP = \frac{1}{N} \sum AP_K \quad (d)$$

Utilizando o conjunto de dados personalizado de uma única classe, artêmia, com um conjunto de dados de 700 imagens que foi treinado com o modelo “yolov5l.pt” disponibilizado pelos criadores da rede YOLO. Foi treinado em 80 épocas usando Pytorch em uma GPU Tesla T4. Os resultados dos três principais termos discutidos acima foram plotados em um gráfico como mostrado abaixo.

Figura 3 – Precisão, Recall, mAP@0.5, mAP@0.5:0.95, de 80 épocas de artêmias plotadas no gráfico.



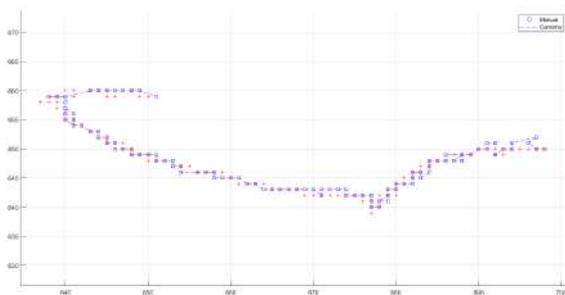
Fonte: Autoria própria, 2022.

Os resultados definem claramente que em um intervalo de 0-35 épocas a precisão estava crescendo e mAP estava aumentando e enquanto chegando às últimas épocas o mAP estava quase estável para o conjunto de dados de

amostra, a precisão ou mAP é aproximadamente 64,39 %. Assim, fomos capazes de treinar um modelo mediano yolo personalizado que pode ser usado para detecção. Uma vez que o treinamento foi concluído, os arquivos de pesos são usados para detecção de objetos em vídeos. O arquivo de vídeo é dado como entrada e cada um de seus frames é processado e nosso detector de objeto treinado e uma vez feita as informações da posição são passadas para o algoritmo de rastreamento de objetos realizando o mapeamento das posições dos objetos detectados no tempo para definir seus respectivos caminhos no vídeo.

Para fins de comparação, apresentamos a seguir uma mapa de um mesmo vídeo marcado de forma manual frame a frame comparada ao mesmo vídeo utilizando a detecção de objetos baseada no YOLO.

Figura 4 – Mapeamento utilizando YOLO para detecção ampliado



Fonte: Autoria própria, 2022.

A Figura 4, mostra os mesmos pontos de mapeamento em um mesmo vídeo e é perceptível a atuação do algoritmo de tracking desenvolvido durante a pesquisa, nas imagens é possível ver o trajeto percorrido pelas artêmias em água durante o vídeo marcado em azul manualmente e um trajeto detectado automaticamente por meio do algoritmo desenvolvido. Em especial é possível avaliar visualmente o comportamento do algoritmo em relação ao real e seu desempenho foi satisfatório.

Conclusões

O reconhecimento de objetos em imagem com base em um padrão específico e métricas de distância permite resolver as tarefas de classificação e busca de objetos. São considerados os métodos de reconhecimento, classificação e busca dos objetos de diversas formas para comparar o desempenho dos mesmos.

Métodos são usados para reconhecer rostos em imagens usando várias métricas de distância. Nos métodos de reconhecimento, a precisão do resultado depende da escolha correta do padrão e da métrica de distância. Nos

métodos de reconhecimento de imagem, as seguintes métricas de distância são usadas: distância euclidiana, distância euclidiana média, média ao quadrado. Usar apenas uma das métricas nos métodos de reconhecimento nem sempre dá um resultado preciso.

Para efeitos comparativos o ganho de inserção de um framework classificador como o YOLO, permite ao usuário ter maior clareza dos dados de saída devido a forma com que o classificador incrementa a percepção e influencia diretamente na obtenção de resultados mais relevantes utilizando uma ascendente tecnologia do mercado, as Redes Neurais Convolucionais.

Em relação aos resultados finais observados é perceptível que o treinamento desenvolvido durante a pesquisa não foi suficiente para classificar com uma efetividade de mercado, esse comportamento foi diagnosticado como limitação de captura de informações por parte do hardware utilizado. Entretanto os Resultados obtidos apontam para um algoritmo de trackeamento com efetividade em efetuar o mapeamento de entidades em um vídeo e seu desempenho pode ser aperfeiçoado em projetos futuros.

Agradecimento

Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), pelo apoio financeiro. Universidade Federal de Itajubá (UNIFEI), pelo apoio e disponibilização de ambiente de pesquisa.

Referências

- 1 - SOLOMON, C.; BRECKON, T. Fundamentals of Digital Image Processing: A practical approach with examples in Matlab. [S.l.]: John Wiley & Sons, 2011. 8
- 2 - NAGENDRAN, A. et al. Recognition and tracking moving objects using a moving camera in complex scenes. International Journal of Computer Science, Engineering and Applications, Citeseer, v. 4, n. 2, p. 31, 2014.
- 3 - REDMON, J. et al. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. [S.l.: s.n.], 2016. p. 779–788. 9
- 4 - FREDRIKSSON, T. et al. Data labeling: An empirical investigation into industrial challenges and mitigation strategies. In: SPRINGER. International Conference on Product-Focused Software Process Improvement. [S.l.], 2020. p. 202–216. 10
- 5 - CHANDOLA, V.; BANERJEE, A.; KUMAR, V. Anomaly detection: A survey. ACM computing surveys (CSUR), ACM New York, NY, USA, v. 41, n. 3, p. 1–58, 2009. 10
- 6 - DANIELSSON, P.-E. Euclidean distance mapping. Computer Graphics and image processing, Elsevier, v. 14, n. 3, p. 227–248, 1980. 15
- 7 - ZAMAN, F. H. K. et al. Visual-based motorcycle detection using you only look once (yolo) deep network. In: IOP PUBLISHING. IOP Conference Series: Materials Science and Engineering. [S.l.], 2021. v. 1051, n. 1, p. 012004. 16