

A UTILIZAÇÃO DE FERRAMENTAS DE ANOTAÇÃO DE IMAGENS NO TREINAMENTO DE ALGORITMOS DE VISÃO COMPUTACIONAL

Elisete Cristina Vidotti da Rocha¹; Cláudia Simões Pinto da Cunha Lima²

¹ Estudante de Desenvolvimento de Sistemas - Escola Técnica - SENAI Cimatec; Iniciação Tecnológica Jr, SENAI Cimatec; Salvador-BA; liz.vidotti@gmail.com

² Doutoranda em Modelagem Computacional e Tecnologia Industrial - SENAI CIMATEC; Salvador-BA; claudia.lima@fieb.org.br

RESUMO

Anotação de dados gera oportunidade para o aprendizado de máquina. O presente estudo tem como objetivo apresentar as experiências vivenciadas, com a utilização de ferramentas de anotação de imagens, no treinamento de algoritmos de visão computacional. Foi desenvolvido um projeto de Iniciação Tecnológica Junior, na Escola Técnica do Senai CIMATEC, para a detecção e segmentação de objetos com a utilização das ferramentas CVAT e Coco Annotator. Com essas ferramentas foram aplicadas bounding boxes (caixas de seleção) e feitas segmentações (contornos) em objetos identificados nos frames (imagens). Estes programas de código aberto aceleraram o processo de anotação de imagens e vídeos no treinamento de sistema de inteligência artificial de um negócio no ramo alimentício.

PALAVRAS-CHAVE: CVAT, Coco Annotator, Visão Computacional, Inteligência Artificial.

1. INTRODUÇÃO

A Inteligência Artificial é uma ciência que busca utilizar as máquinas para a realização de atividades humanas, que vão desde a automatização de atividades repetitivas e manuais, além de atividades de análise e tomada de decisão. O termo Inteligência Artificial surgiu no final da década de 1950 com o professor John McCarthy, numa conferência científica¹. Há uma tendência do uso de IA nas mais diversas áreas, a exemplo da medicina, no reconhecimento facial usado em sistemas de segurança, no marketing, no direito e em outras áreas. Com a IA é possível mapear tendências, comportamentos, demandas e oportunidades, gerando engajamento e impulsionando os negócios¹.

A inteligência artificial compreende diversas tecnologias, a exemplo do *Machine Learning*, *Deep Learning*, Processamento de Linguagem Natural e Visão Computacional, que auxiliam no processo de aprendizagem da máquina, resolvendo problemas mais complexos e trazendo resultados mais consistentes². Já a Visão computacional é uma área da Inteligência Artificial focada em detectar e extrair informações a partir de imagens e vídeos. Podemos perceber a utilização da Visão Computacional na Medicina, com diagnósticos automatizados a partir da análise de radiografias, ressonâncias e ultrassonografias, e até na tecnologia dos carros autônomos, que são capazes de identificar as sinalizações do trânsito, e outros veículos que estejam transitando nas ruas³.

São diversas as possibilidades de resolução de problemas que utilizam da visão computacional, como automatização de tarefas, classificação de imagens, detecção e segmentação de objetos⁴. A classificação de imagens é a forma mais comum. É possível classificar números, caracteres e imagens simples. A detecção de objetos pode ser considerada uma evolução da classificação de imagens, pois ela é capaz de detectar mais de um objeto em uma mesma imagem. Para exemplificar, considerando uma área onde se percebe vários animais como gatos, cachorros e passarinhos, é possível classificar quais objetos (animais) são gatos e quais objetos são cachorros. Essa detecção é feita com o uso de caixas de seleção sobre cada objeto que deseja-se identificar. Com a segmentação de objetos, a atividade é ainda mais precisa, eles (os objetos) são contornados seguindo sua forma exata, ao invés de caixas de seleção⁴.

O processo de entendimento, análise e predição dos dados é possível graças ao *Machine Learning*, ou Aprendizado de Máquina, que criam os modelos de análise dos dados e reconhecimento de padrões, a partir da aplicação de algoritmos. Dessa forma, é possível que os modelos criados sejam capazes de analisar um grande volume de dados complexos e entregar os resultados com rapidez e eficiência⁵.

Para garantir o sucesso do modelo de *Machine Learning*, os algoritmos aplicados devem ser bem treinados. O treinamento consiste na alimentação de entradas e saídas desejadas de dados. Quase sempre o volume de dados que a aplicação final recebe é muito grande, por isso, é importante que durante o treinamento também possa receber bom volume de dados, aumentando as chances de detecção, análise e predições mais consistentes. Se queremos, por exemplo, que um *e-mail* diferencie as mensagens legítimas de mensagem de *spam*, devemos utilizar vários tipos de *e-mail*, para que o aprendizado da máquina seja mais eficiente⁶.

Esses algoritmos de *Machine Learning* possuem dois tipos principais: aprendizado supervisionado e aprendizado não-supervisionado. No aprendizado supervisionado, os dados são rotulados e o modelo deve aprender quais são as categorias que devem ser identificadas. No aprendizado não-supervisionado, não há um objetivo específico para o modelo e os dados são não-rotulados. O modelo deve ser capaz de reconhecer padrões e fazer relacionamentos entre os dados⁵.

Em resumo, para o treinamento de algoritmos de *Machine Learning* é necessário fazer a anotação de dados. E para o aprendizado supervisionado, a anotação de dados é feita com a marcação dos objetos a serem identificados e analisados pelo modelo. As marcações podem ser feitas com *bounding box* ou com polígonos delimitando as bordas. Como é necessário um grande volume de dados, em geral são utilizadas ferramentas de anotação de imagens, a exemplo do CVAT (*Computer Vision Annotation Tool*) e *Coco Annotator*, softwares de código aberto baseados na *web* que realizam o processo de forma mais rápida.

Diante do contexto apresentado, o objetivo deste estudo é apresentar a importância da utilização de ferramentas de anotação de imagens para o treinamento de algoritmos de visão computacional através de um relato de experiência da utilização das ferramentas CVAT e *Coco Annotator*.

2. METODOLOGIA

Esse projeto de Iniciação Tecnológica Jr, desenvolvido na Escola Técnica do SENAI Cimatec, na cidade de Salvador, Bahia, foi desenvolvida por uma equipe de cinco bolsistas. Foram realizadas reuniões para definição dos critérios para a classificação das imagens e compreensão da utilização das ferramentas de anotação de imagens. Para a execução das atividades, foram usadas duas ferramentas: CVAT (*Computer Vision Annotation Tool*) e *Coco Annotator*. O CVAT é um programa de código aberto que acelera o processo de anotação de imagens e vídeos para uso em computadores de treinamento de algoritmos de visão computacional⁷.

Com o CVAT é possível anotar imagens em formas de caixas, polígonos, polilinhas e pontos. Já a ferramenta *Coco Annotator*, que também é um programa de código aberto e baseado na *web*, cria dados de treinamento para localização de imagens e detecção de objetos, e fornece muitos recursos, como rotular um segmento de imagem, rastrear instâncias de objetos, rotular objetos parcialmente cobertos, e exportar as anotações no formato *Coco*. As imagens podem ser anotadas em formas de polígonos, curvas e pontos⁸.

No processo de anotação de imagens foram trabalhadas a detecção e segmentação de objetos. A detecção de objetos foi utilizada na ferramenta CVAT, com a marcação das *bounding boxes*, ou caixas de seleção para detecção dos objetos que deveriam ser reconhecidos nas imagens disponibilizadas. Os objetos foram divididos em classes específicas. Em uma mesma imagem, podiam ser marcados objetos pertencentes a diferentes classes. Já na ferramenta *Coco Annotator*, trabalhou-se com a segmentação de objetos, onde eram marcados os *endpoints*, ou pontos. O conjunto dos pontos pertenciam a uma classe específica, que delineava a estrutura do corpo humano.

3. RESULTADOS E DISCUSSÃO

Para que uma empresa possa usufruir intensamente de um sistema de inteligência artificial, obtendo resultados que possam colaborar com a otimização dos processos, melhoria na produção e logística, se faz necessário investir tempo no treinamento do sistema. O aprendizado da máquina deve ser acompanhado de um processo de treinamento eficiente, para que a análise e predição dos dados seja feita de maneira rápida e eficaz. Assim, a tarefa exige que sejam feitos com um grande volume de dados, dificultando o processo de anotação de imagens. Foram utilizadas as ferramentas CVAT e *Coco Annotator*, que visam à eficiência e versatilidade no processo de anotação de imagens⁸.

Na ferramenta CVAT foram usados principalmente as caixas de seleção para a identificação de objetos e definição de pontos para as estruturas que necessitam de maior precisão de reconhecimento. As anotações foram feitas em frames de vídeo de uma empresa do ramo alimentício (frigorífico), e as caixas de seleção serviam para a marcação, por exemplo, da cabeça de boi, e os pontos serviam para marcação dos olhos e do focinho do animal.

Na ferramenta *Coco Annotator* foram usados os *endpoints*, utilizados nos pontos-chave (olhos, nariz, orelhas, ombros, cotovelos, pulsos, quadris, joelhos e tornozelos) do corpo humano. Esse tipo de anotação é mais complexa e demanda mais tempo quando comparado ao uso do CVAT.

Ambas ferramentas facilitam o trabalho em equipe, sendo possível atribuir grupos de imagens, ou *jobs*, a cada bolsista do projeto, de forma a estimular a colaboração e trabalho em grupo. É possível utilizá-las através do navegador de *internet*, evitando que sejam instalados programas específicos no computador.

Porém, pode ocorrer lentidão do processamento de imagens, a depender da velocidade de conexão da *internet*. A facilidade no processo de anotação vem acompanhada de um processo mais lento de teste, com a correção manual. Uma vantagem dessas duas ferramentas é a anotação de objetos que estejam parcialmente cobertos por outro objeto, além da anotação de objetos em imagens com pouca iluminação, ou resolução mais baixa.

4. CONSIDERAÇÕES FINAIS

Para o presente estudo, onde foi possível experimentar duas ferramentas de anotação de dados, observou-se a importância de se ter a preparação do ambiente para a coleta de dados, a exemplo do posicionamento das câmeras, onde são captados os vídeos e depois gerados os frames. As imagens devem ter boa resolução, para que não haja dificuldades de reconhecimento dos objetos por parte dos profissionais responsáveis pela anotação dos dados, e esse fato não influencie na eficiência do algoritmo de visão computacional.

A experiência vivenciada no projeto permite afirmar a importância do envolvimento da equipe e o treinamento para a realização das atividades. Observou-se que a cada inclusão de novo objeto para marcação, a equipe precisava de tempo para entender as características e como proceder com a anotação. No entanto, a possibilidade de utilização de atalhos do teclado torna o trabalho mais rápido. Segundo relatos da equipe, nas reuniões de alinhamentos, a maior dificuldade apresentada foi seguir o critério para a marcação dos pontos-chave do corpo humano e outro fator é o tamanho reduzido dos pontos, que mesmo utilizando da ferramenta de *zoom*, precisa de habilidade para a seleção e movimentação destes.

Já com a ferramenta CVAT, o processo de aprendizado da equipe do projeto foi mais rápido. A marcação das imagens utilizando as caixas retangulares (chamadas *bounding boxes*) é mais rápida e com maior área para cliques e arraste das caixas.

5. REFERÊNCIAS

¹TOTVS, Equipe. **O que é Inteligência artificial? Como funciona, exemplos e aplicações**. TOTVS, 2019. Disponível em: <https://www.totvs.com/blog/inovacoes/o-que-e-inteligencia-artificial/>. Acesso em 10/04/2021.

²ROCKCONTENT, Redator. **Inteligência Artificial: entenda o que é e como ela funciona**. ROCKCONTENT, 2019. Disponível em: <https://rockcontent.com/br/blog/inteligencia-artificial/>. Acesso em 10/04/2021.

³BRANDIZZI, Loreane. **Visão Computacional? O que é? Como funciona?: Uma breve explicação sobre os sistemas artificiais que permitem que computadores "enxerguem"**. SERPRO, 2020. Disponível em: <https://www.serpro.gov.br/menu/noticias/noticias-2020/o-que-eh-visao-computacional>. Acesso em 25/03/2021.

⁴SOUZA, Maria Fernanda. **Carreira em Visão Computacional: como se preparar para trabalhar na área: O que é Visão Computacional, quais os principais tipos de aplicação, o que você precisa aprender e referências para guiar seus estudos**. Programaria, 2020. Disponível em: <https://www.programaria.org/carreira-em-visao-computacional-como-se-preparar-para-trabalhar-na-area/>. Acesso em 25/03/2021.

⁵WALTRICK, Camila. **Machine Learning — O que é, tipos de aprendizagem de máquina, algoritmos e aplicações**. Camila Waltrick, 2020. Disponível em: <https://medium.com/camilawaltrick/introducao-machine-learning-o-que-e-tipos-de-aprendizado-de-maquina-445dcfb708f0>. Acesso em 10/04/2021.

⁶SOLVIMM. **A Etapa de Treinamento de Algoritmos de Machine Learning**. SOLVIMM, 2018. Disponível em: <https://solvimm.com/blog/a-etapa-de-treinamento-de-algoritmos-de-machine-learning/>. Acesso em 10/04/2021.

⁷SEKACHEV, Boris; MANOVICH, Nikita; ZHAVORONKOV, Andrey. **New Computer Vision Tool Accelerates Annotation of Digital Images and Video: We are pleased to present a new open source program called Computer Vision Annotation Tool**. Intel. Disponível em: <https://www.intel.com/content/www/us/en/artificial-intelligence/posts/introducing-cvat.html>. Acesso em 29/03/2021.

⁸BROOKS, Justin. **Coco Annotator**. Github, 2019. Disponível em: <https://github.com/jsbroks/coco-annotator/wiki>. Acesso em 29/03/2021.