Conceitos teóricos

YoloV3

YOLO é um algoritmo de deteção de objetos em tempo real que está atualmente na sua terceira versão. É um algoritmo bastante rápido e preciso comparado aos concorrentes. YOLO funciona através de uma aplicação de uma rede neural simples para a imagem. Esta rede divide a imagem entre regiões e prevê caixas delimitadoras e probabilidades em cada região. Estas caixas delimitadoras estão ponderadas pelas probabilidades previstas [13]. Contém uma versão Tiny para máquinas menos potentes, como por exemplo, CPUs de gama U da Intel ou da AMD. O algoritmo do YOLOv3/Darknet está disponível no GitHub de pjreddie em: <https://github.com/pjreddie/darknet> .

No contexto do projeto, YOLOv3/Darknet é utilizado para a deteção e classificação de objetos em tempo real.

Darknet

CUDA

CUDA (Compute United Device Architecture) é uma plataforma de computação paralela e uma API (Application Programming Interface) criada pela Nvidia, lançada a 23 de junho de 2007. CUDA é suportado por GPUs que sejam da Nvidia, a sua licença pertence também à Nvidia e é do tipo GPGPU (General Purpose Processing computing on Graphics Processing Units). CUDA permite aos engenheiros e desenvolvedores de software para usar uma GPU que seja compatível com CUDA para uso de processamento geral sobre a forma de GPGPU. CUDA é uma camada de software que dá acesso direto ao conjunto de instruções virtuais da GPU e elementos de computação paralela, para a execução de kernels. É desenhada para trabalhar com linguagens de programação como C, C++ ou Fortran, esta acessibilidade permite ser mais fácil programar em paralelo usando os recursos da GPU [12].

(Falta meter o uso específico do CUDA no projeto)

Machine Learning

Machine Learning é um tipo de inteligência artificial que permite que aplicações de software possam ser bastantes precisas na previsão de resultados, mesmo que não estejam programadas, de forma específica, para isso. Consiste na execução de algoritmos que criam de forma automática modelos de representação de conhecimento com base num conjunto de dados. A ideia deste tipo de inteligência artificial assenta no treino das máquinas, dando-lhas acesso a dados históricos, uma ou mais medidas de desempenho, deixando assim o algoritmo “aprender”, ou seja, de modo iterativo o modelo de representação de conhecimento de modo que este melhore o seu desempenho. Após o treino, o modelo tem potencial para efetuar previsões de qualidade em situações posteriores e que estejam relacionadas com padrões históricos/anteriores. Este método de análise de dados estão bastante enraizado no nosso dia-a-dia, algumas usos de Machine Learning, temos, por exemplo, as pesquisas na Web, as recomendações da Amazon ou o Google Tradutor [11].

Deep Learning

Deep Learning é um tipo de machine learning que permite treinar PCs para realizar tarefas como se atuasse como um ser humano, permite o reconhecimento de imagem, previsões e reconhecimento da fala. Ao contrário de alguns machine learnings, que organiza os dados para serem executados através de equações predefinidas, o Deep Learning configura parâmetros básicos sobre os dados e treina o PC para aprender sozinho através do reconhecimento de padrões em várias camadas de processamento [1]. Alguns exemplos de softwares que utilizam Deep Learning, temos o caso da Cortana, da Microsoft, que está disponível no Windows 10, porém, não funciona em Portugal ou a Siri, da Apple, disponíveis nos aparelhos da companhia. Quer a Cortana, quer a Siri, são uma assistente inteligente que ajuda o utilizador a realizar algumas tarefas no Sistema Operativo.

OpenCV

OpenCV (Open Source Computer Vision Library) é uma biblioteca de funções para programação focada para visão computacional em tempo real, criada originalmente pela Intel em junho de 2000, escrita em C++. A sua biblioteca é cross-platform com interfaces de C++, Java e Python, que funcionam em Linux, macOS, Windows, iOS e Android. Dentro destas escolhas, nós escolhemos a interface de Python para realizar o projeto, sobre a licença gratuita e open source Apache 2. Desde 2011 que o OpenCV traz funcionalidades de aceleração de GPU para operações em tempo real. O OpenCV inclui áreas de aplicação onde inclui sistema de reconhecimento facial, realidade aumentada, captura de movimento, reconhecimento gestual, interação máquina-pessoa, classificador de Naive-Bayes, algoritmo do k nearest neighbor, redes neurais artificiais, deteção de objeto, que é a área onde o OpenCV é aplicado no projeto, entre várias outas áreas [2]. O OpenCV contém um GitHub oficial, que pode ser acedido em: <https://github.com/opencv/opencv> , documentação online e tutoriais online disponibilizada de forma gratuita e cursos de aprendizagem de IA através de pagamento. Contém parcerias com várias empresas, tais como a Intel, o Microsoft Azure ou a Google[3].

PostgreSQL

PostgreSQL é um software gratuito e open source de gerenciamento de base de dados relacional que enfatiza a extensibilidade e a complacência em SQL, lançado inicialmente a 8 de julho de 1996, é escrito na linguagem C. Está disponível nos sistemas macOS, onde é a base de dados pré-definida para macOS Server, Windows, Linux, FreeBSD e OpenBSD. Contém transações com as propriedades ACID (Atomicidade, Consistência, Isolamento e Durabilidade), visualizações atualizadas de forma automática, vistas materializadas, triggers, chaves estrangeiras e stored procedures/procedimentos armazenados (isto soa mal). Está desenhado para lidar com uma variedade de cargas de trabalho, desde máquinas simples até Data Warehouses ou Web Services com um número forte de usuários em simultâneo [6].

A escolha desta base de dados para a realização do projeto deve-se a três razões: A existência de documentação oficial para PostgreSQL, acessível de forma gratuita em: <https://www.postgresql.org/docs/> , o Microsoft Azure ter suporte total com PostgreSQL permitindo assim o armazenamento de dados da deteção dos objetos para consultas posteriores e, por fim, pelo facto de já estarmos familiarizados com esta base de dados, graças à realização de trabalhos em Unidades Curriculares anteriores.

Microsoft Azure

O Microsoft Azure é um serviço de cloud computing, criado pela Microsoft para construir, testar, implementar e, aplicações e serviços administrativos, através de data centers fornecidos pela Microsoft. Providencia software como serviço (Os chamados SaaS – Software as a Service), plataforma como serviço (Os chamados PaaS – Platform as a Service) e infraestruturas como serviço (IaaS – Infrastucture as a Service). Suporta várias linguagens de programação, frameworks, incluindo softwares e sistemas third-party quer software e sistemas da própria Microsoft. O Azure foi apresentado pela primeira vez a 27 de outubro de 2008 e lançado oficialmente em fevereiro de 2010. O serviço funciona no Windows e no Linux, a licença é fechada para plataforma, porém é open source para SDKs (Software Development Kit) para clientes [4].

Dentro da plataforma, uma conta Azure tem acesso a mais de 200 produtos que são fornecidos pelo Azure, dependendo da categoria de trabalho. Dentro destas categorias, são fornecidas ferramentas para, referindo alguns exemplos: Ambiente de Trabalho Virtual do Windows, análise de dados, armazenamento, Blockchain, computação, programação, gestão, IA + Machine Learning, Internet das Coisas, multimédia, segurança ou web [5].

No contexto do projeto, o Azure vai servir-nos para um processamento rápido de captura de imagem ou de vídeo para o reconhecimento de objetos em tempo real e para a interface web que é desenvolvida numa máquina virtual. Ora, havendo estes requisitos para a realização do projeto, a criação de uma máquina virtual com o sistema operativo Ubuntu, que irá correr a interface web, é feita no Azure, assim como a base de dados do Azure para PostgreSQL para o armazenamento de dados e consultas posteriores.

O Azure só foi utilizado na parte final do projeto, devido a problemas com as contas do Azure fornecidas pela UAL, entre nunca se conseguir ter acesso à conta ou a licença terminar de forma abrupta. Dito isto, a construção da base de dados e criação da interface web foram inicialmente realizadas nos PCs dos elementos do grupo e depois portado para o Azure.

Flask

Flask é uma micro framework web escrita em Python, lançada a 1 de abril de 2010. A sua licença é BSD. É classificada desta maneira porque não requere ferramentas ou bibliotecas particulares, mantendo um núcleo simples, porém, extensível. Flask não possui uma camada de abstração de base de dados, nem de validação de formato, nem de outros tipos de componentes onde bibliotecas third-parties providenciam funções comuns. Sites como o Pinterest ou o Linkedin usam esta framework[7]. O Flask contém um GitHub que pode ser acedido em: <https://github.com/pallets/flask> e a sua documentação é gratuita, disponível em: <https://flask.palletsprojects.com/en/2.0.x/>

O uso de Flask no projeto está relacionada com a construção da página web onde apresenta os resultados e a classificação da deteção de objetos da imagem ou do vídeo representado.

Python

Python é uma linguagem de programação de alto-nível, multi-paradigma, ou seja, suporta vários paradigmas de programação, lançada em fevereiro de 1991. É uma das linguagens de programação mais populares [8]. A sua documentação está disponível livremente em: <https://docs.python.org/3/> . A escolha desta linguagem no projeto, deve-se a vários motivos: Para começar, a versão do Python que foi utilizada no projeto foi a versão 3.9.4. É uma linguagem na qual já trabalhamos bastante com ela ao longo do curso, que estamos familiarizados com ela para construir o programa requerido, para a utilização do Flask, que é uma framework de Python para páginas web, para fazer os testes do YOLOv3 e do Darknet para a deteção de alvos e também por ser uma linguagem que se dá bem com a linguagem SQL, focada para criação de base de dados, através da biblioteca SQLAlchemy do Python, que será expandida no ponto a seguir. Todas estas vantagens fizeram que escolhêssemos Python como a linguagem de programação a ser utilizada no projeto.

SQLAlchemy

SQLAlchemy é uma toolkit open-source de SQL e mapeamento objeto-relacional (ORM – Object-relacional Mapping em inglês) para linguagem Python, lançada sobre a licença MIT a 14 de fevereiro de 2006[9]. Providencia um conjunto completo de padrões de persistência de nível empresarial bem conhecidos, desenhados para eficiência e alta performance no acesso à base de dados, adaptada numa simples linguagem e dominante como é Python [10]. Empresas como o Reddit ou a DropBox utilizam SQLAlchemy. Contém um GitHub oficial, disponível em: <https://github.com/sqlalchemy/sqlalchemy> e a sua documentação está disponível em: <https://www.sqlalchemy.org/library.html> .

A escolha da utilização no SQLAlchemy no projeto deve-se à facilidade de conectar o código Python do algoritmo de deteção e a página web com a base de dados, Postgresql, escrita em SQL, na mesma linguagem.

[1] <https://www.sas.com/pt_br/insights/analytics/deep-learning.html> , Deep Learning, O que é qual sua importância?, consultado a 18/5/2021

[2] <https://en.wikipedia.org/wiki/OpenCV> , OpenCV, consultado a 18/5/2021

[3] <https://opencv.org/> , consultado a 18/5/2021

[4] <https://en.wikipedia.org/wiki/Microsoft_Azure> , Microsoft Azure, consultado a 18/5/2021

[5] <https://azure.microsoft.com/pt-pt/services/> , Produtos do Azure, consultado a 18/5/2021

[6] <https://en.wikipedia.org/wiki/PostgreSQL> , PostgreSQL, consultado a 18/5/2021

[7] <https://en.wikipedia.org/wiki/Flask_(web_framework)> ,Flask, consultado a 19/5/2021

[8] <https://en.wikipedia.org/wiki/Python_(programming_language)> . Python, consultado a 19/5/2021

[9] <https://en.wikipedia.org/wiki/SQLAlchemy> , SQLAlchemy, consultado a 19/5/2021

[10] <https://www.sqlalchemy.org/> , The Python SQL Toolkit and Object Relational Mapper, consultado a 19/5/2021

[11] <https://www.ccg.pt/machine-learning-o-que-e/> , Machine learning: o que é e para que serve?, publicado a 13 de dezembro de 2017, consultado a 25/5/2021.

[12] <https://en.wikipedia.org/wiki/CUDA> , CUDA, consultado a 25/5/2021.

[13] <https://pjreddie.com/darknet/yolo/> , YOLO: Real-Time Object Detection, How It Works, consultado a 25/5/2021.