Conceitos teóricos

YoloV3 e Darknet

O cerne do projeto era detetar e classificar objetos em tempo real e era necessário usarmos um algoritmo que o fizesse de forma rápida e precisa. Já nos estava especificado no enunciado que enumerava o âmbito do projeto, a existência do YOLO.

O YOLO, uma sigla para You Only Look Once (Apenas Olhas Uma Vez), é o algoritmo mais rápido, preciso e estável disponível para processar o que nós precisávamos. E não só, também existia uma vasta biblioteca de tutoriais e documentação que nos iria ajudar imenso.

O YOLO funciona através de uma aplicação de uma rede neural simples para a imagem, chamada de Darknet. Cada nó dessa rede tem apenas um objetivo, comparar uma região da imagem com um conjunto de dados de um dataset. Dividindo a imagem em regiões e caixas delimitadores facilita a previsão das probabilidades de o objeto estar na imagem ou não [13].

Em termos de performance, existe uma versão mais fraca do YOLO, chamada de Tiny YOLO. Esta é usada para máquinas menos potentes e portáteis, desde CPUs de portátil, até Raspberry Pis [14].

O Darknet enquadra-se nesta questão como quase um adendo e gestor do algoritmo do YOLO. No fundo, o YOLO são blocos de algoritmo a ser corridos pelo Darknet, que depois de os organizar e gerir, monta-nos o seu output, em forma de imagem.

No entanto, por cada instrução em que o Darknet é chamado, o Yolo/Darknet fazem output de apenas uma única imagem de cada vez. E para processarmos streams de vídeo das câmaras, era necessário uma biblioteca que, a partir do input de um ficheiro de vídeo, processasse esta rede neural várias vezes.

O algoritmo do YOLOv3/Darknet está disponível no GitHub de pjreddie em: <https://github.com/pjreddie/darknet>.

OpenCV

Open Source Computer Vision Library é o nome completo de OpenCV, que é uma biblioteca de funções para visão computacional em tempo real, criada originalmente pela Intel em junho de 2000. Este campo de estudo, a visão computacional, trata de desenvolver métodos para robots e computadores poderem processar imagens e vídeos como informação. Sendo escrita em C++, satisfaz a necessidade de celeridade e de portabilidade destas funções, ajudando no desenvolvimento moderno de software para drones e robôs que têm como objetivo a análise de imagens da forma mais rápida e leve possível.

Dito isto, o OpenCV tem interfaces para Java e Python, não limitando-se apenas a C++, que funcionam em todos os sistemas operativos principais do mercado. A nossa escolha foi a interface para Python, sob a licença gratuita de código aberto Apache 2.

Tudo o que o OpenCV é capaz de fazer inclui áreas de aplicação com sistemas 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, sendo a última a de interesse especial para o nosso projeto[2].

No entanto, era necessário que este processamento fosse o mais rápido possível, portanto necessitamos não só do processamento do nosso processador, mas também do nosso processador gráfico. E desde 2011 que o OpenCV traz funcionalidades de aceleração de GPU para operações em tempo real. Agora só era necessário uma plataforma para ligarmos o OpenCV com uma placa gráfica.

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].

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, informalmente conhecido como ML, é um tipo de inteligência artificial que permite que aplicações de software possam ser mais precisas e eficazes no processamento e na previsão de resultados, apesar de nenhum código específico tenha sido escrito para processar os dados. Isto significa que em vez de adaptarmos o código a cada um dos cenários possíveis para controlar o fluxo do processamento, simplesmente damos um objetivo ao programa que será o guia para ver se as interações estão mais perto do objetivo ou não.

Em termos técnicos, ML nada é mais que a interação de algoritmos que criam de forma automática modelos de representação de conhecimento com base num conjunto de dados. Sendo estes modelos de representação o que o algoritmo já aprendeu e melhorou por si mesmo.

E antes de poder ser usado, um algoritmo é deixado a processar um conjunto de dados de treino para afinar a sua precisão. Após o treino, o modelo tem potencial para efetuar previsões de qualidade em situações posteriores e que estejam relacionadas com padrões anteriores. Basicamente informação nova, mas do mesmo tipo que a que foi usada para treino. Seja texto, rostos humanos ou animais, um dos exemplos que pode ser um tipo de dado analisado por algoritmos ML.

No entanto, ML está limitado na sua única função e apenas faz uma coisa bem. Isto significa que para outras funções mais complexas, ML simplesmente não chega. Precisamos de algo mais complexo.

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 algoritmos ML, que organizam 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.

PostgreSQL

PostgreSQL é um software gratuito e open source de gestão de base de dados relacional, feito para ser compatível com SQL agnóstico. Como outros softwares de base de dados, o PostgreSQL efectua transações com as propriedades ACID (Atomicidade, Consistência, Isolamento e Durabilidade), visualizações atualizadas de forma automática, vistas materializadas, triggers, chaves estrangeiras e stored procedures. 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]. Para nós, apenas nos interessou três características: o facto de já termos usado o software anteriormente na universidade, a vasta documentação disponível online e o suporte completo dado pelo Microsoft Azure a esta tecnologia.

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.

Bootstrap

Bootstrap é uma framework front-end de CSS gratuita e open-source lançada pela Bootstrap Core Team a 19 de agosto de 2011. É responsiva, ótima para desenvolvimento de páginas web em dispositivos móveis ou diferentes tipos de janelas. Vai neste momento na versão 5.0.1, é escrita em HTML, CSS, Less, Sass e JavaScript e está sobre a licença MIT License. Bootstrap contém templates de design de CSS e de JavaScript para tipografia, formas, botões, navegação e outras componentes de interface[15]. Contém um GitHub oficial em: <https://github.com/twbs/bootstrap> , a sua instalação está disponível em: <https://getbootstrap.com/docs/5.0/getting-started/download/> e contém documentação oficial em: <https://getbootstrap.com/docs/5.0/getting-started/introduction/>.

No projeto, o uso de Bootstrap serve para a Interface Web ser compatível com qualquer dispositivo e de qualquer tamanho (por exemplo telemóveis), sem prejudicar o conteúdo disponível no site.

CSS, JavaScript e HTML

HTML, CSS e JavaScript são linguagens de programação viradas para programação web, que estão bastante conectadas entre si. Começando por HTML, significa HyperText Markup Language é a linguagem standard para documentos mostrados em um browser de Internet, criada em 1993. É assistido por tecnologias como o CSS (Cascade Style Sheet) e linguagens de script como o JavaScript. Os browsers de Internet recebem documentos HTML a partir de um servidor web ou a partir de um lugar local e faz renderização dos documentos em multimédia na página web [16].

Já CSS (Cascade Style Sheet), é uma linguagem de folha de estilos, lançada em 1996, usada para mostrar um documento escrito em uma linguagem de marcos como HTML. É uma das linguagens principais da World Wide Web ao lado de HTML e JavaScript. CSS é projetado para possibilitar a separação da apresentação e do conteúdo, incluindo layouts, fontes ou cores [17].

JavaScript é uma linguagem de programação que está em conformidade com a especificação ECMAScript, lançada inicialmente a 4 de dezembro de 1995. É uma linguagem de alto nível, normalmente compilada em tempo real e multi-paradigma. Contém orientação a objetos e funções de primeira classe. Ao lado de HTML e CSS, JavaScript é uma das linguagens mais usadas na World Wide Web, enquanto linguagem de multi-paradigma, JavaScript suporta paradigmas de programação imperativa, funcional e orientado por eventos. Contém APIs para trabalho de texto, estruturas de dados, datas ou DOM (Document Object Model, Modelo de Objeto de Documentos em português) [18].

No projeto estas três linguagens foram essenciais para a construção da Interface Web requerida no enunciado do projeto.

Ubuntu

[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.

[14] <https://www.pyimagesearch.com/2020/01/27/yolo-and-tiny-yolo-object-detection-on-the-raspberry-pi-and-movidius-ncs/>, YOLO and Tiny-YOLO object detection on the Raspberry Pi and Movidius NCS, consultado a 25/5/2021

[15] <https://en.wikipedia.org/wiki/Bootstrap_(front-end_framework)>, Bootstrap (front-end framework), consultado a 31/5/2021

[16] <https://en.wikipedia.org/wiki/HTML> , HTML, consultado a 1/6/2021

[17] <https://en.wikipedia.org/wiki/CSS>, CSS, consultado a 1/6/2021

[18] <https://en.wikipedia.org/wiki/JavaScript> , JavaScript, consultado a 1/6/2021