



UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”

Faculdade de Ciências

Bacharelado em Sistemas de Informação

Luis Guilherme Silva Rodrigues

**EDIÇÃO DE METADADOS DE IMAGENS UTILIZANDO INTELIGÊNCIA
ARTIFICIAL**

Bauru – SP

2019

Luis Guilherme Silva Rodrigues

**EDIÇÃO DE METADADOS DE IMAGENS UTILIZANDO INTELIGÊNCIA
ARTIFICIAL**

Trabalho de conclusão de curso apresentado ao curso de Bacharelado em Sistemas de Informação da Universidade Estadual Paulista "Júlio de Mesquita Filho", como requisito parcial para obtenção de título de Bacharel em Sistemas de Informação.

Orientador: Prof. Dr. José Remo Ferreira Brega

Bauru – SP

2019

RESUMO

Nas últimas décadas, o armazenamento de dados foi altamente facilitado, o custo para armazenar grandes quantidades de dados foi reduzido e está acessível para a maioria da sociedade. Com um simples smartphone é possível criar textos, imagens, vídeos e qualquer outro tipo de mídia, compartilhar o conteúdo na internet e armazenar os arquivos em vários outros dispositivos. Resgatar estes arquivos para serem acessíveis em um momento futuro pode ser uma tarefa difícil se não estiverem organizados. O tipo de mídia abordado neste trabalho será a imagem. Assim como o armazenamento e criação de dados, a inteligência artificial também evoluiu e pode auxiliar na organização de imagens, reconhecendo padrões e classificando imagens automaticamente.

Palavras-chave: Fotografias. Imagens. Python. Inteligência Artificial.

Reconhecimento Facial. Detecção de Objetos. Transferência de Aprendizado.

ABSTRACT

In recent decades, data storage has been greatly facilitated, the cost of storing large amounts of data has been reduced and is affordable for most of society. With a simple smartphone you can create text, images, videos and any other media, share content on the internet and store the files on various other devices. Rescuing these files for future access can be a daunting task if they are not organized. The type of media covered in this paper will be the image. Like data storage and creation, artificial intelligence has also evolved and can help organizing images, recognizing patterns and classifying images automatically.

Keywords: Photographs. Images. Python Artificial intelligence. Facial Recognition. Object Detection. Transfer Learning.

SUMÁRIO

1 INTRODUÇÃO	8
1.1 Detalhamento do Problema	9
1.2 Soluções Existentes	10
1.2.1 ACDSsee	10
1.2.2 Google Photos.....	11
1.2.3 Photos do Windows.....	13
1.2.4 Cloud Vision	13
1.3 Metadados.....	14
1.3.1 Padrão EXIF	15
1.3.2 Padrão IPTC.....	16
2 PROJETO DO SOFTWARE	17
2.1 Especificação	17
2.1.1 Diagrama de classes	18
2.1.2 Diagramas de atividades	18
2.2 Tecnologias Propostas	21
2.2.1 Plataforma do Software	21
2.2.2 Reconhecimento Facial.....	21
2.2.3 Detecção de Objetos	22
2.2.4 Classificação e Transferência de Aprendizado	22
3 IMPLEMENTAÇÃO.....	24
3.1 Python e PyQt4.....	24
3.2 face_recognition.....	24
3.3 YOLOv3	25
3.4 Keras	25
4 USO DA FERRAMENTA	26
4.1 Metadados.....	26

4.2 Pessoas.....	27
4.3 Objetos	29
4.4 Classificação	31
4.5 Resgate de Imagens.....	34
5 CONCLUSÃO	35
5.1 Desafios Encontrados.....	35
5.2 Aspectos Positivos	36
5.3 Melhorias Futuras	36
REFERÊNCIAS BIBLIOGRÁFICAS.....	38

LISTA DE ILUSTRAÇÕES

FIGURAS

Figura 1 – Reconhecimento facial do software ACDSee	11
Figura 2 – Reconhecimento facial do Google Photos.....	12
Figura 3 – Recurso de pesquisa do Google Photos.....	12
Figura 4 - Aplicativo Photos do Windows	13
Figura 5 – Cloud Vision e detecção de objetos por API.....	14
Figura 6 – Exemplo de metadados EXIF.....	15
Figura 7 – Classes envolvidas no problema	18
Figura 8 – Ações do usuário: Editar fotos e ensinar o sistema	19
Figura 9 – Funções dos modelos de inteligência artificial	20
Figura 10 – Uma pessoa, um cachorro e um carro na foto.....	22
Figura 11 – Imagens que envolvem o mesmo domínio: Números.....	23
Figura 12 – Visualização de metadados de uma imagem.....	26
Figura 13 – Adicionando uma nova pessoa.....	27
Figura 14 - Resultado da função de reconhecimento facial	28
Figura 15 – Assertividade da detecção em diferentes ângulos do rosto	28
Figura 16 – Encontrando objetos em uma pasta.....	29
Figura 17 - Teste com objetos diversos	30
Figura 18 – Seleção das pastas que o modelo utilizará para treino	31
Figura 19 – Processamento de uma pasta com classes aleatórias	32
Figura 20 – Botão para visualizar imagens em forma de olho	34
Figura 21 – Todas as imagens que possuem determinado objeto	34

QUADROS

Quadro 1 – Cronograma	17
Quadro 2 – Matriz de confusão com pouco treinamento.....	32
Quadro 3 – Matriz de confusão: Cores de cabelo	33
Quadro 4 – Matriz de confusão: fotos de ensaios fotográficos	33

1 INTRODUÇÃO

Desde o surgimento das primeiras máquinas fotográficas digitais houve a necessidade de armazenar os arquivos de imagens, enviar essas imagens para o computador ou para algum serviço *web*, liberando assim o espaço no cartão de memória do dispositivo que tirou a foto.

No cenário atual há a um número crescente de smartphones, que contam com câmeras digitais cujas tecnologias proporcionam aos usuários maior facilidade seja na obtenção quanto na qualidade de fotografias, o que colabora para o aumento do acervo pessoal dos usuários.

A maneira mais fácil de armazenar uma foto é copiá-la para uma pasta de backup no computador, onde o armazenamento é maior. Entretanto, não é a forma mais conveniente quando se quer encontrar a imagem rapidamente.

Considerando as ferramentas de armazenamento disponíveis, é possível fazer mais com imagens além de colocá-las em pastas: O usuário, ao inserir uma imagem, pode definir um conjunto ou álbum e até definir etiquetas para determinada imagem. Ambas as opções se tornam metadados para o arquivo e será possível resgatar a respectiva imagem e outras semelhantes. Esta atividade é feita para toda imagem nova e mesmo que seja um conjunto de fotos, o usuário precisa selecionar as imagens individualmente.

É importante pontuar que os arquivos de imagens digitais estão presentes nos mais diversos meios de comunicação como no comércio eletrônico, em redes sociais, em aplicativos de mensagens instantâneas e também na própria galeria do usuário, pois este, além de consumir e salvar algumas imagens interessantes da web em seu dispositivo, também é gerador de novas mídias com sua câmera pessoal, culminando num acervo de difícil organização manual.

Deste modo, ressalta-se a importância do presente trabalho que propõe como ferramenta um gerenciador de imagens desenvolvido a partir de ferramentas de inteligência artificial como uma solução alternativa para a problemática identificada.

O documento do trabalho está estruturado resumidamente por seções de identificação do problema, apresentação de soluções existentes, descrição e implementação do projeto, uso e resultados da ferramenta e conclusões.

Com a presença desse tipo de arquivo em diversos assuntos e até mesmo a presença de diferentes conjuntos de imagens na câmera do usuário, surge a

necessidade de organização caso o mesmo tenha o desejo de manter esses arquivos para encontrá-los futuramente.

Assim, imagens que demandavam tempo para serem resgatadas ou ficavam arquivadas sem nenhum destino passam a ser acessíveis.

Apresentar uma forma sistêmica de organizar esses arquivos pode servir como referência para a organização e categorização de outros tipos de arquivos. O formato de arquivo PDF, por exemplo, possui metadados que podem ser adicionados ao arquivo e ajudam a identificar o documento sem precisar abri-los (ADOBE ACROBAT, 2017).

É provável que o problema se torne cada vez mais recorrente devido ao crescimento das redes sociais e o compartilhamento de fotos tanto para pessoas quanto para empresas que lidam com imagens diretamente.

1.1 Detalhamento do Problema

Será apresentado neste trabalho uma alternativa para a visualização nas imagens armazenadas em um computador e uma interface que auxilie na classificação de novas imagens para a estrutura já existente. O *software* pretende tomar decisões automaticamente para que poupe o usuário de avaliar as imagens individualmente e que ele gaste menos tempo com a tarefa de organizar fotos.

Para isso, recursos de aprendizado de máquina e redes neurais serão utilizados.

Manter fotos pessoais de forma organizada demanda tempo e não fazer isso pode acarretar praticamente na perda de fotos por não saber onde a foto ficou armazenada.

Estimativas apontam que mais de 1 trilhão de fotos digitais são tiradas por ano no mundo e os números são ainda maiores quando se trata de compartilhamento.

Em um dispositivo eletrônico também existem outros formatos de arquivo além de imagens, existem arquivos de vídeo, músicas e documentos. Todos são passíveis de serem organizados, entretanto, será abordado a organização de imagens.

A natureza da necessidade de organização sempre ocorreu, até mesmo antes das mídias digitais. Até hoje papéis são utilizados para documentar processos e informações sobre pessoas ou fatos, bibliotecas precisam organizar livros de acordo com categorias e até mesmo em mercados e lojas, produtos são fisicamente organizados de forma que itens similares fiquem próximos um do outro.

A principal ideia é criar uma interação entre o usuário e o modelo que ele poderá personalizar de acordo com suas necessidades e utilizar o *software* como uma ferramenta de aprendizado para esta tecnologia, enquanto também executa uma tarefa comum que é arquivar ou visualizar fotos pelo computador.

Introduzir tecnologias novas, mais especificamente da inteligência artificial, em ferramentas de uso cotidiano ainda é um desafio pois exige ferramentas de desenvolvimento e conhecimento que não são tradicionais no mercado (ainda). Trazer uma abordagem inovadora para solucionar o problema especificado no começo dessa sessão pode impactar na tarefa considerada comum, que é a solução sem o uso de aprendizado de máquina, logo, é preciso pensar em como transformar um assunto complexo em algo aceitável para o usuário final.

1.2 Soluções Existentes

É crescente o número de pesquisas acadêmicas sobre inteligência artificial e também um alto crescimento no número de startups e no uso da mesma por empresas já consolidadas no mercado.

Além do grande conjunto de imagens existentes e arquivos em geral, também é uma realidade a visão computacional e o reconhecimento de imagens. Já está disponível para o nível do usuário aplicações que reconhecem superfícies pela câmera e originam realidade aumentada, existe também reconhecimento facial que identificam pessoas e a possibilidade de pesquisar exclusivamente por imagens.

As aplicações com inteligência artificial supracitadas melhoram a experiência do usuário, auxilia em tarefas que são repetitivas e até cria novas funcionalidades.

1.2.1 ACDSee

Este *software* é conhecido por ser um organizador de fotos e possui a melhor avaliação entre os *softwares* concorrentes. Possui uma interface personalizável, para visualizar as fotos e suas respectivas *tags* e miniaturas da forma que o usuário preferir.

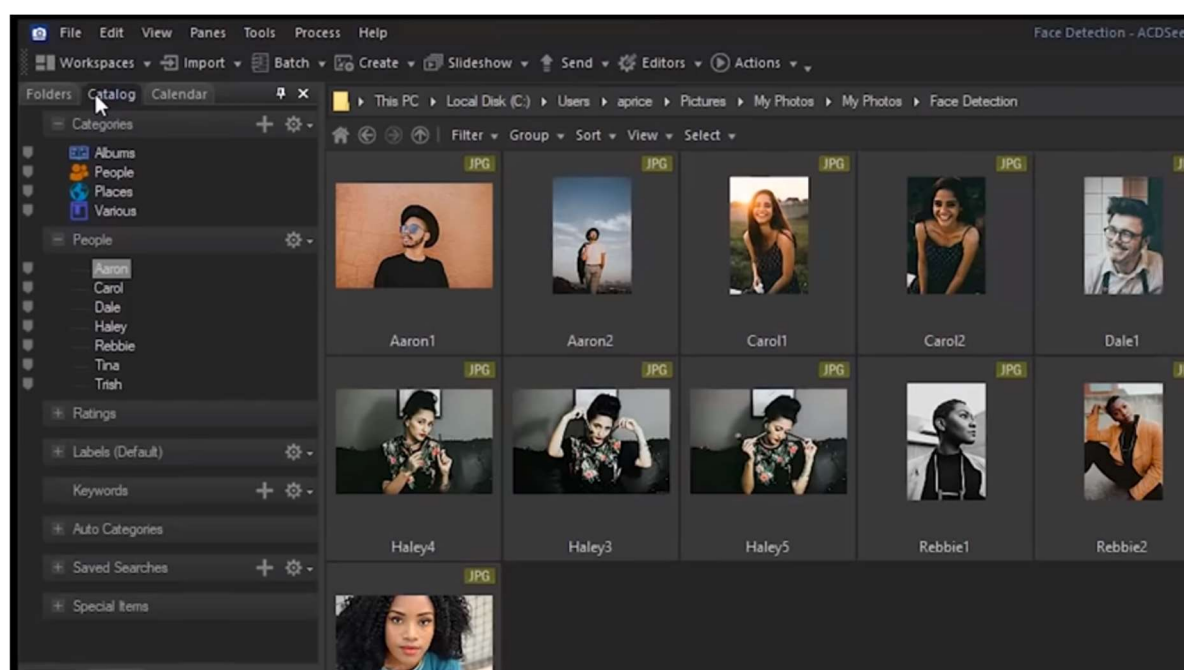
Algumas das funcionalidades são:

- a) Suporte para qualquer tipo de arquivo de imagem;
- b) Ferramentas de conversão de formato, redimensionamento e edição de imagens;

c) Organização por *tags*, incluindo reconhecimento facial e *tags* automáticas nas fotos.

A imagem seguinte é uma captura de tela obtida de um vídeo explicativo do *software*. À esquerda há uma lista com pessoas conhecidas, como Aaron, Carol e Dale, e no centro há imagens contendo as pessoas da lista. Uma mesma pessoa pode aparecer em mais de uma foto e com a lista é possível encontrar essas fotos de determinada pessoa facilmente.

Figura 1 – Reconhecimento facial do *software* ACDSee



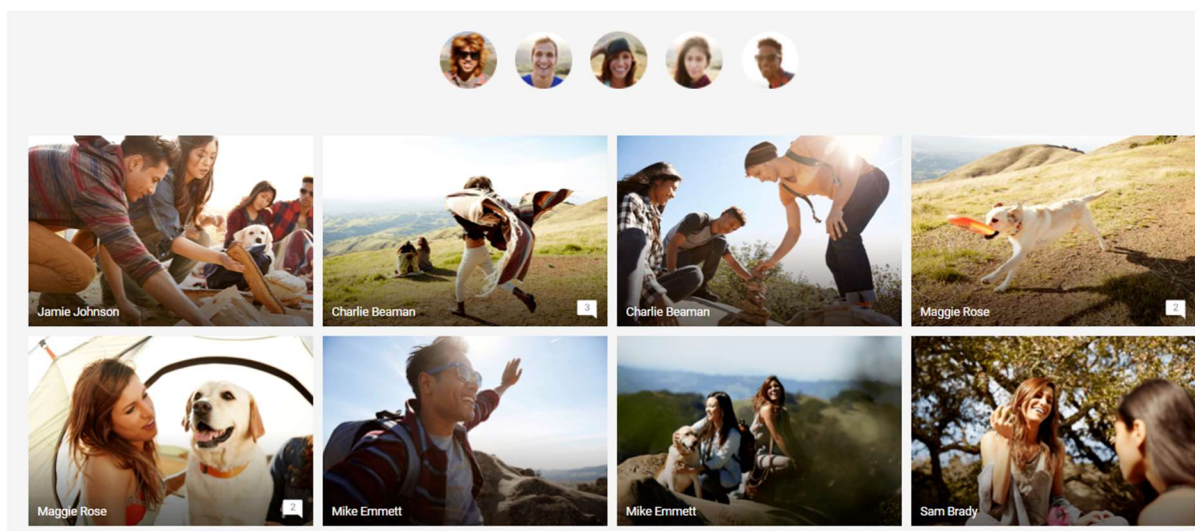
Fonte: ACDSee (2019).

1.2.2 Google Photos

É um aplicativo altamente integrado com os outros serviços do Google. Possui armazenamento ilimitado de fotos na nuvem e é muito utilizado em celulares como visualizador de imagens.

A capacidade de organização não é limitada somente por pastas e nas versões atuais, o sistema categoriza imagens por pessoas. Assim como no *software* anterior, o usuário precisa identificar a pessoa na foto somente uma vez e após essa atribuição do nome, é possível exibir fotos desta pessoa digitando os nomes desejados na barra de pesquisa.

Figura 2 – Reconhecimento facial do Google Photos



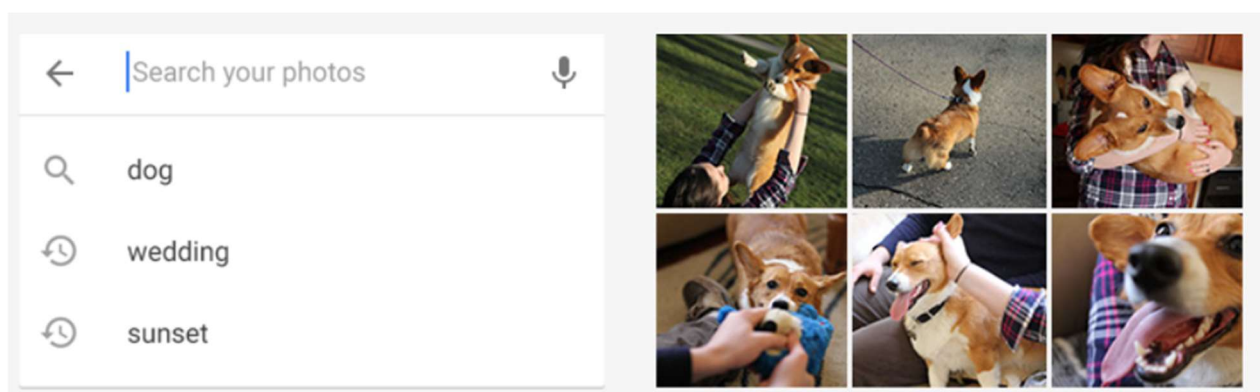
Fonte: Google Photos (2019)

A barra de pesquisa por texto deste aplicativo não só pode encontrar pessoas pelo nome como também pelo nome de um animal, por objetos, ou lugares, assim, as imagens são filtradas, aparecendo somente fotos em que o animal esteja presente, por exemplo.

Como é dito na própria página de apresentação do Google Photos (2019), não é preciso usar *tags* nas fotos para fazer a pesquisa de coisas e objetos, entretanto, reduz o nível de compatibilidade com outros aplicativos, pois essa informação estará disponível somente no Google Photos.

De acordo com Rita El Khoury (2019), há a possibilidade de corrigir quando as pessoas que foram marcadas automaticamente não são elas de fato, deixando o usuário editar manualmente. Entretanto não é possível editar outros metadados.

Figura 3 – Recurso de pesquisa do Google Photos



Fonte: Google Photos (2019)

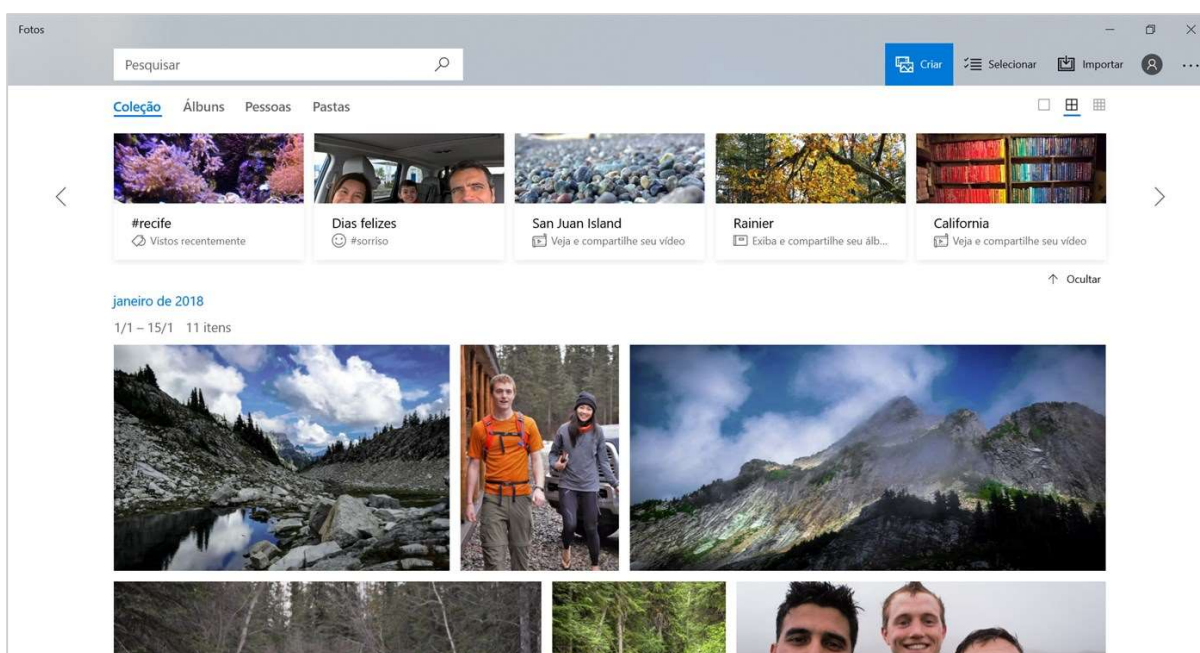
1.2.3 Photos do Windows

Photos é um aplicativo padrão do sistema Windows 10 para visualização e edição de fotografias e vídeos.

O aplicativo consegue identificar as faces das pessoas presentes nas fotos e as agrupam em uma aba específica. Quando quiser ver todas as fotos com determinada pessoa, basta selecioná-la na lista.

Além disso, são sugeridas palavras para a pesquisa, como Recife e Sorriso, pois é identificado o que há na imagem e também pesquisa pela localização da imagem, como no aplicativo Google Photos.

Figura 4 – Aplicativo Photos do Windows



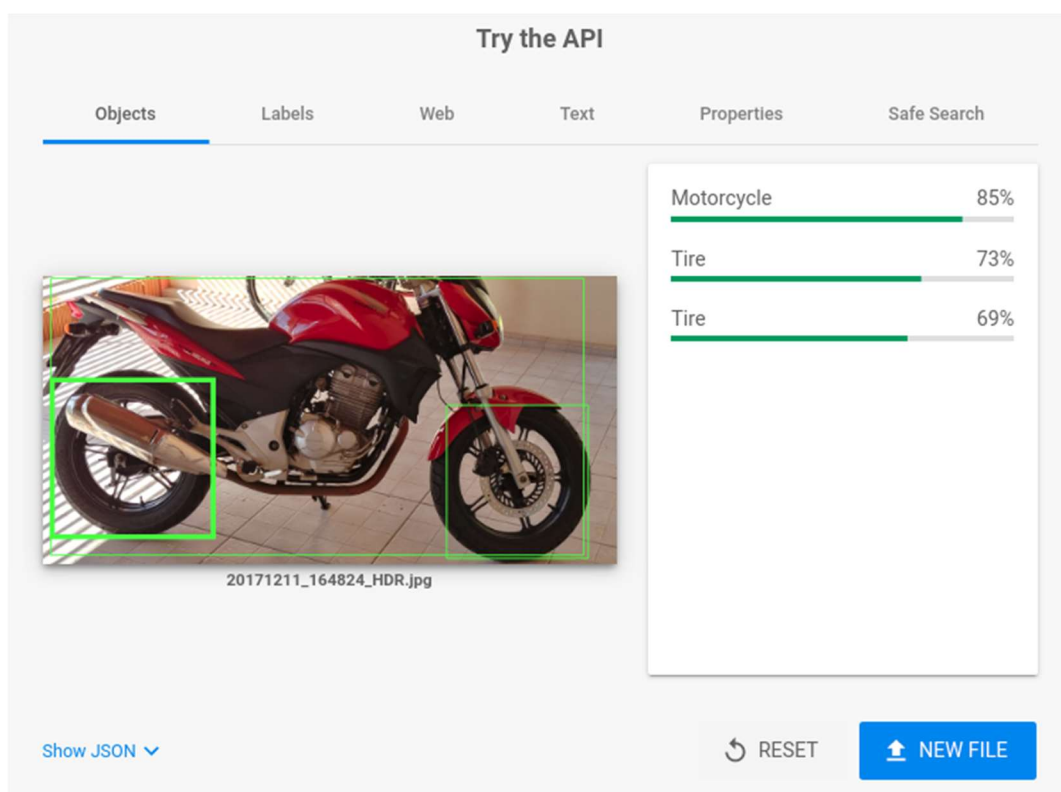
Fonte: Rahul (2019)

1.2.4 Cloud Vision

É um serviço comercializado pela de API para a classificação de imagens, feita a chamada com uma imagem, o serviço retorna a lista de objetos identificados.

Por ser somente uma API, não é possível utilizá-la diretamente para armazenar, organizar ou recuperar fotos sem o desenvolvimento de uma aplicação que incorpore esse serviço.

Figura 5 – Cloud Vision e detecção de objetos por API



Fonte: Elaborado pelo autor (2019)

Também existem outras alternativas, com a existência desse tipo de serviço, o *software* proposto neste trabalho pode ser estruturado de uma forma que utilize estes tipos de serviços em que existem modelos pré-treinados prontos para prever novas imagens.

A API permite que outros sistemas possam usufruir de inteligência artificial sem a necessidade de criar o próprio modelo, pode ser utilizada até mesmo no projeto proposto como um meio de obter informações sobre a imagem.

1.3 Metadados

Metadados em arquivos de imagens mantêm informações sobre o conteúdo da fotografia, direitos autorais, detalhes técnicos da câmera, termos para pesquisas, pessoas retratadas e outras informações (IPTC, 2019).

Estas informações podem até implicar em mudanças na relevância de sites que contêm esse tipo de imagem (GOOGLE WEBMASTERS, 2014).

A importância da utilização de padrões é a facilidade de comunicação com diferentes softwares, mantendo o usuário apto a usufruir das informações em diversas plataformas.

1.3.1 Padrão EXIF

É um padrão de metadados mantido pela Associação de Indústrias de Tecnologia da Informação e Eletrônicos do Japão e pela Associação de Produtos de Câmera e Imagem (CAMERA & IMAGING PRODUCTS ASSOCIATION, 2010).

Dentre os atributos mais utilizados estão o tamanho e as dimensões a imagem, data e hora e informações da câmera utilizada.

No entanto, imagens com informações sobre a localização do dispositivo utilizado são menos frequentes e até retiradas por questões de segurança e privacidade (MOREY, 2018).

Figura 6 – Exemplo de metadados EXIF

File Name	_N5A3412.CR2	
Sidecar Files	xmp	
File Path	21	→
Dimensions	8688 x 5792	
Cropped	8688 x 5792	→
Date Time Original	2018-02-21 13:49:28	→
Date Time Digitized	2018-02-21 13:49:28	
Date Time	2018-02-21 13:49:28	
Exposure	1/640 sec at f / 2.8	
Focal Length	180 mm	
Exposure Bias	0 EV	
ISO Speed Rating	ISO 320	→
Flash	Did not fire	
Exposure Program	Aperture priority	
Metering Mode	Pattern	
Make	Canon	
Model	Canon EOS 5DS R	
		→
Lens	180mm	→

Fonte: Nell (2018)

1.3.2 Padrão IPCT

International Press Telecommunication Council Information Interchange Model é uma estrutura criada por um consórcio entre setores de notícias e ainda é o padrão mais utilizado.

Entre os diversos campos existentes nesse padrão estão (IPTC, 2017):

- Autor;
- Aviso de Direitos Autorais;
- Cidade;
- Data de Criação;
- Descrição;
- E-mail;
- Endereço;
- Gênero Intelectual;
- Telefone;
- Título;
- País;
- Palavras-Chave;
- Província/Estado.

Enquanto o padrão EXIF possui majoritariamente atributos numéricos ou pré-determinados, no padrão IPTC existem atributos livres, em que o usuário pode inserir textos que desejar, como em “Descrição” e em “Palavras-chave”.

Existem atributos úteis e exclusivos em cada um dos padrões, logo, ambos podem ser utilizados até mesmo em conjunto para o mesmo fim: organização e facilidade da pesquisa por imagens, como explica Zeman (2018).

2 PROJETO DO SOFTWARE

O *software* proposto é uma alternativa aos *softwares* apresentados na seção 1.2, sobre soluções existentes, por possuir semelhanças na maneira de visualizar as imagens e também na introdução de técnicas de aprendizado de máquina como ferramenta auxiliadora, como no Google Photos e Photos do Windows.

No capítulo 1, foi analisado o contexto em que o software pode ser inserido. Após isso, como apresentado neste capítulo, foi definido quais funções poderão existir, antes de começar a implementação, incluindo um detalhamento de cada tecnologia necessária para determinadas funções.

Também foi definido uma sequência de atividades a serem executadas a cada mês, para manter o controle de todas as atividades necessárias para a conclusão do trabalho. O conjunto de tarefas abrange a criação ou aperfeiçoamento de funções do *software* ou atividades de pesquisa e documentação.

Quadro 1 – Cronograma

Março	Definição do assunto tratado
Abril	Detalhamento do problema escolhido
Maio	Elaboração dos primeiros diagramas
Junho	Criação de um protótipo
Julho	Funções básicas do software
Agosto	Reconhecimento facial
Setembro	Deteção de objetos
Outubro	Classificação de imagens
Novembro	Documentação

Fonte: Elaborado pelo autor (2019).

2.1 Especificação

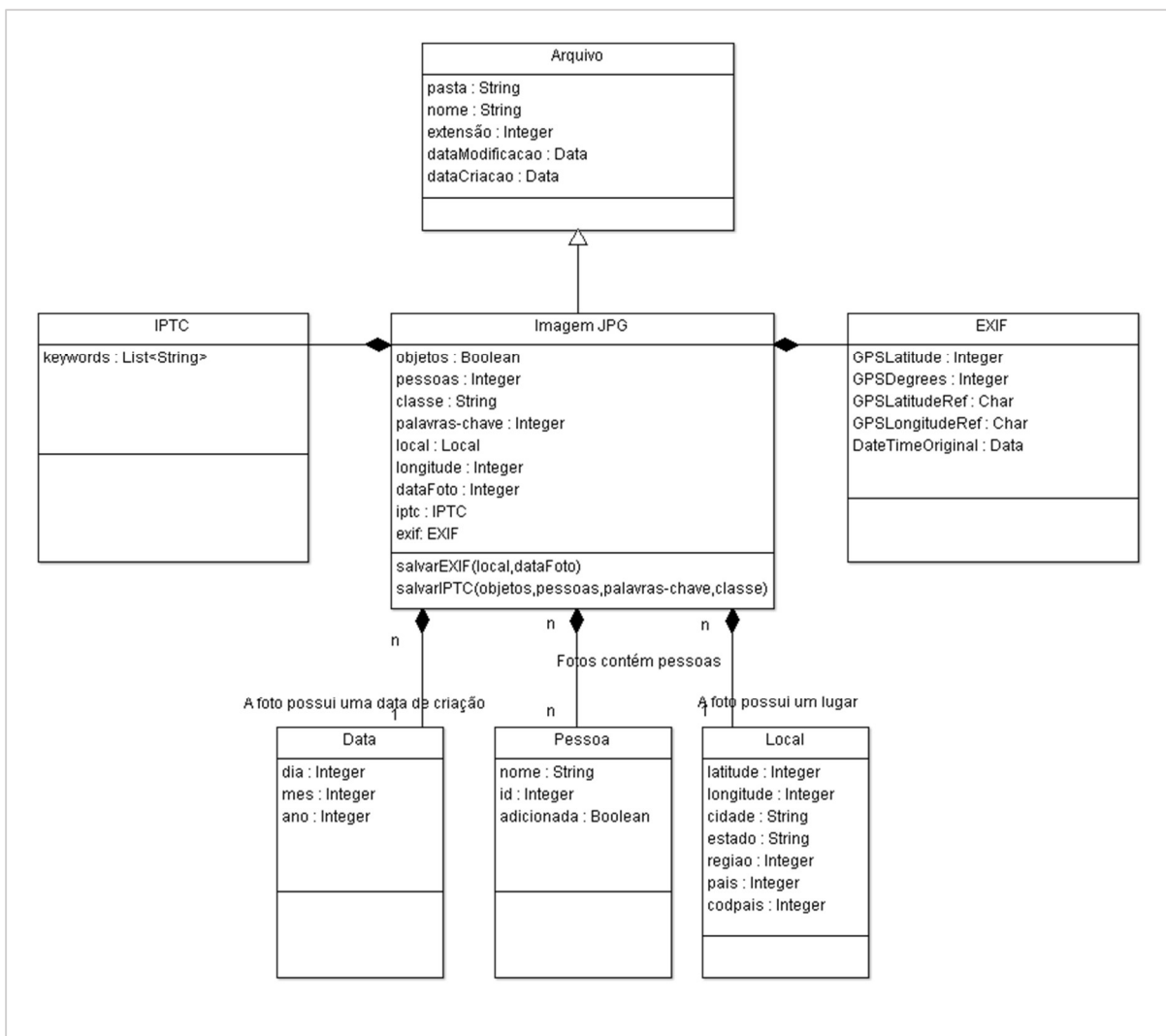
Os diagramas seguintes apresentam o escopo do projeto e os comportamentos tanto do usuário quanto do sistema.

Primeiramente é apresentado um diagrama de classes e posteriormente dois diagramas de atividade. Estes diagramas UML são úteis para apresentar o problema tanto para usuários, que podem utilizar o sistema, quanto para desenvolvedores.

2.1.1 Diagrama de classes

No diagrama de classes da Figura 7 é apresentado os principais atributos trabalhados no projeto. Com ele, é possível obter a base para o código do software e entender qual a delimitação do projeto.

Figura 7 – Classes envolvidas no problema

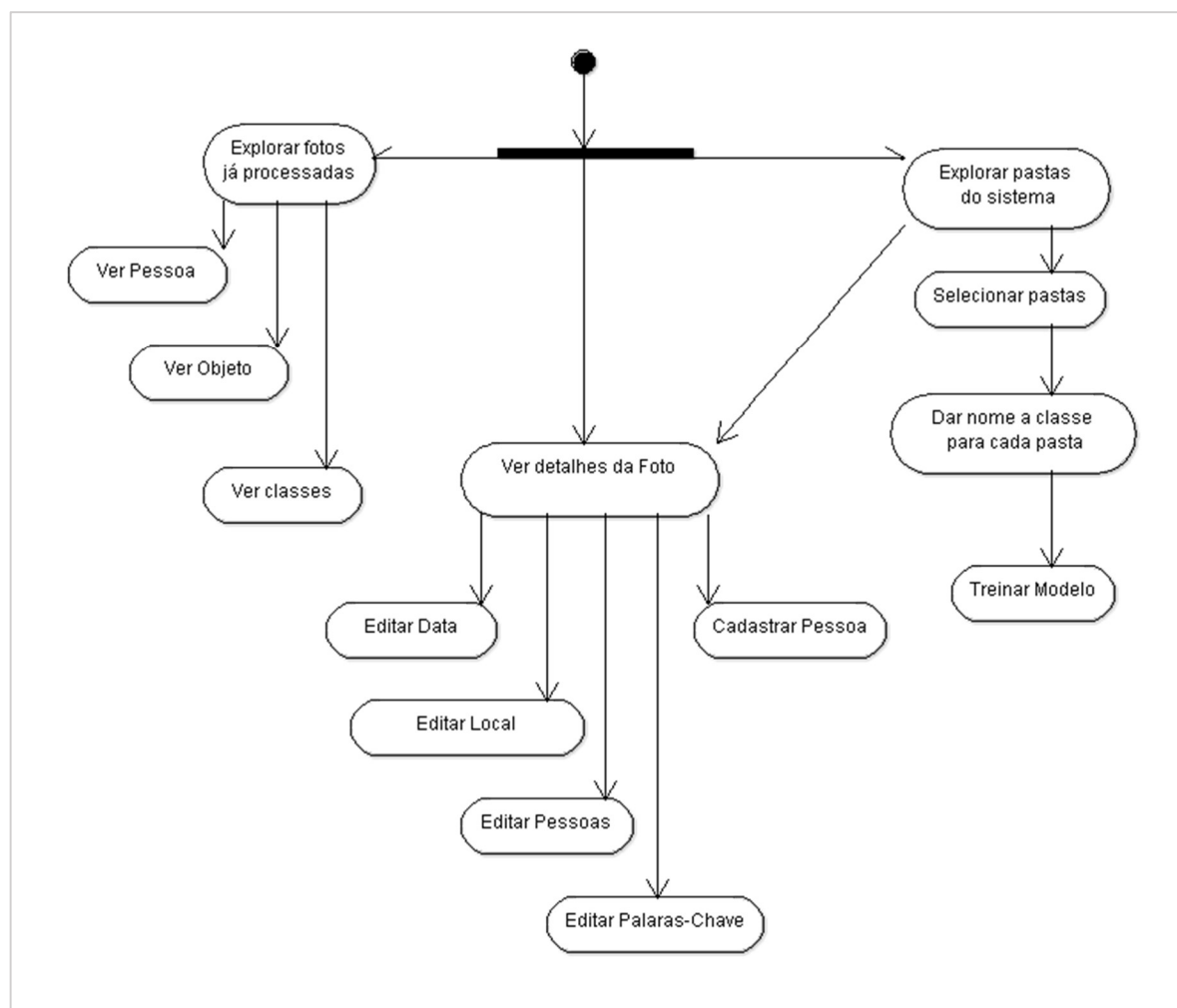


Fonte: Elaborado pelo Autor (2019)

2.1.2 Diagramas de atividades

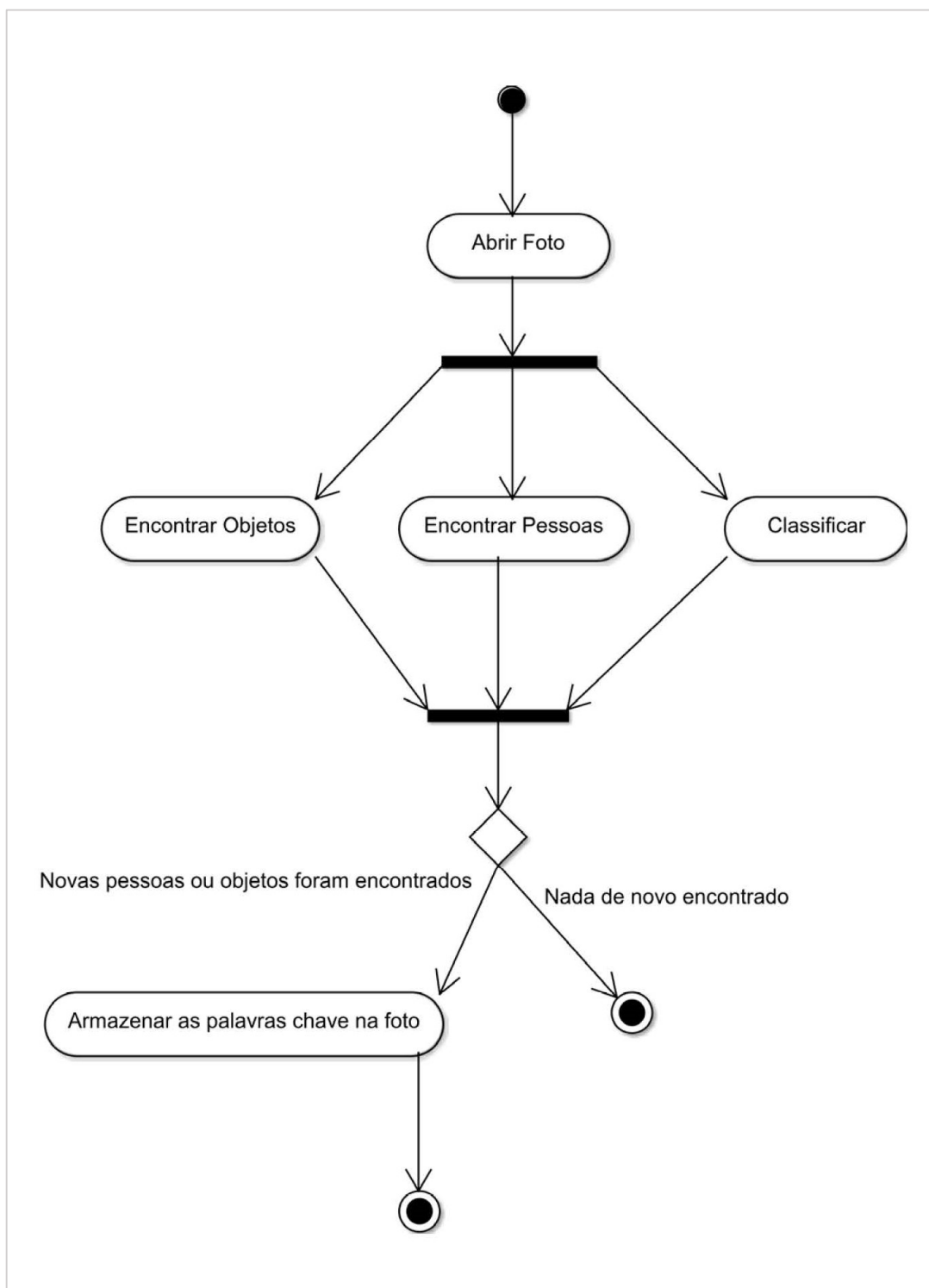
O diagrama na Figura 8 apresenta as funcionalidades que o usuário pode acessar, como explorar imagens, editar metadados, e ensinar o sistema (Cadastrar Pessoa e Treinar Modelo).

Figura 8 – Ações do usuário: Editar fotos e ensinar o sistema



Fonte: Elaborado pelo autor (2019)

Com as ações do usuário, apresentadas na Figura 8, entre elas “Cadastrar Pessoa” e “Treinar Modelo”, o sistema passa a ter também uma parte automatizada. Como é representado mais detalhadamente na Figura 9.

Figura 9 – Funções dos modelos de inteligência artificial

Fonte: Elaborado pelo autor (2019)

2.2 Tecnologias Propostas

Antes da implementação do *software*, foram escolhidas tecnologias para esta aplicação de acordo com o tema proposto e as funções necessárias para o software se comportar de maneira similar às soluções existentes.

Depois de apresentadas as tecnologias, serão listadas quais linguagens de programação e bibliotecas de código fundamentaram e permitiram a criação do software que será apresentado no capítulo 3, que abordará os processos da fase de implementação.

2.2.1 Plataforma do Software

Como modelos de inteligência artificial serão criados e utilizados e estes exigem um nível significativo de processamento, a melhor opção é executar em um computador *desktop*. Descartando versões *mobile* ou para *web*, se mantém praticamente sistemas operacionais como Windows, Linux e Mac.

2.2.2 Reconhecimento Facial

Retratos e fotos com amigos são comuns em álbuns pessoais. Com este recurso é pretendido facilitar a procura de imagens com determinada pessoa e para isso é preciso “marcar” no arquivo que existe uma ou mais pessoas retratadas na foto por meio dos metadados. Após esse procedimento inicial, que se difere do processo manual em que o usuário precisa escrever o nome da pessoa para cada foto, o software emprega o uso de reconhecimento facial que pode facilitar essa tarefa detectando automaticamente as pessoas na foto e marcando na foto as pessoas reconhecidas.

Os principais passos para o funcionamento de sistemas com reconhecimento facial são, de forma geral (NORTON, 2019):

1. Um retrato de uma pessoa, especificamente com o rosto para frente ou quase de perfil, é extraída da foto;
2. O sistema calcula distâncias na geometria do rosto, identificando marcos faciais que distinguem esta face das outras, como uma assinatura facial;

3. Com o resultado do passo anterior, esta assinatura é comparada com outras de faces já existentes no sistema;
4. O reconhecimento é feito caso haja a coincidência da assinatura calculada com alguma cadastrada.

2.2.3 Detecção de Objetos

Em uma foto podem ter diversos objetos cujos nomes são passíveis de serem agregados nos metadados da imagem em formato de texto.

Figura 10 – Uma pessoa, um cachorro e um carro na foto



Fonte: Norcross (2015)

Algoritmos de detecção de objetos localizam instâncias de determinadas classes e podem até mesmo definir contornos de cada objeto na imagem (SZELISKI, 2010, p. 685).

2.2.4 Classificação e Transferência de Aprendizado

Classificação é uma técnica na área de visão computacional que busca prever qual é a classe de determinada imagem.

O treino de um modelo de classificação é realizado a partir de conjuntos categorizados de imagens extraído os padrões para cada classe.

Modelos de transferência de aprendizagem se baseiam na utilização de modelos de classificação treinados que já resolvem algum problema e na reutilização do mesmo para resolver um segundo problema relacionado.

Na Figura 11 é mostrado imagens de um conjunto de dados chamado MNIST (nas duas primeiras linhas) que contém imagens normalizadas de numerais escritos à mão e na última linha existem fotos de números de casas, que são muito parecidos com os dígitos acima. A rede-neural criada para o primeiro problema (MNIST) possui pesos que determinam formas e contornos na imagem que resultam em um número. Com esta rede é possível utilizar parte dos mesmos pesos para resolver o segundo problema, que é reconhecer números de casas.

Figura 11 – Imagens que envolvem o mesmo domínio: Números



Fonte: Ruder (2017)

Com essa tecnologia é possível trazer modelos que têm uma boa performance para a realidade de cada o usuário pois com um modelo que já conhece muitos padrões necessitará de poucas imagens para aprender algo que o usuário desejar.

3 IMPLEMENTAÇÃO

Neste capítulo, são apresentados linguagens, bibliotecas e *frameworks* utilizados para a realização das funções almejadas no capítulo 2, onde foram pensadas na plataforma e as tecnologias de uma forma geral, sem citar bibliotecas.

3.1 Python e PyQt4

Para projetos que envolvem aprendizado de máquina, geralmente são utilizadas linguagens de programação de alto-nível, como Python, para diminuir a quantidade de linhas de código e facilitar o teste.

Para esta linguagem existem bibliotecas pré-compiladas que possuem implementações para estruturas de dados complexas, além de ser compatível com muitos sistemas operacionais.

Para a interface gráfica foi planejado utilizar ferramentas que tenha boa integração com o *back-end* do sistema, sendo assim, foi escolhido um *framework* da mesma linguagem de programação escolhida (Python) para manter o *software* em um mesmo pacote.

Programas PyQt4 podem ser executados em Windows, Linux e Mac OS, além de outras plataformas UNIX. A versão atual de PyQt é a versão 5, que inclui as plataformas Android e iOS. Apesar de as duas plataformas adicionadas em PyQt5, será utilizado PyQt4 pela ausência de tutoriais em PyQt5 e também pela quantidade maior de construção na comunidade de desenvolvedores na versão 4.

3.2 face_recognition

É o pacote responsável por aplicar a técnica de reconhecimento facial. Utiliza como base a biblioteca dlib, que contém um conjunto de algoritmos e ferramentas de aprendizado de máquina.

Alguns dos exemplos presentes na documentação da biblioteca face_recognition (GEITGEY, 2018): Detectar rostos em uma fotografia; desfocar rostos em um vídeo; identificar características faciais específicas em uma fotografia; E identificar e desenhar bordas em volta de cada pessoa em uma foto.

3.3 YOLOv3

You Only Look Once (YOLO) é um sistema de detecção de objetos em tempo real consideravelmente mais rápido que os outros modelos existentes. Atualmente, existem 3 versões, com melhorias a cada versão mais recente, e a utilizada no projeto é a versão 3, publicada em 2016 (FARHADI; REDMON, 2016).

O banco de dados utilizado para treinar o modelo que será agregado no sistema proposto é chamado Open Images, que possui 9.2 milhões de imagens e 15.4 milhões de caixas delimitadoras representando 600 classes (KUZNETSOVA et al., 2018, p. 1).

Os pesos resultantes do treino necessários para carregar o modelo foram obtidos de uma das soluções existentes da competição Google AI Open Images - Object Detection Track que obteve o 19º lugar e disponibilizou o arquivo final para download, facilitando a criação da funcionalidade de detectar objetos em imagens (GOOGLE RESEARCH, 2018).

3.4 Keras

Keras é uma biblioteca de rede neural de código aberto capaz de ser executada acima de outras bibliotecas de aprendizado de máquina como Tensorflow, CNTK e Theano que permite criar protótipos de maneira fácil e rápida através da facilidade de uso, modularidade e extensibilidade.

Com essa API é possível criar modelos de rede neural com diferentes arquiteturas, compilá-los e treiná-los. Dentre as arquiteturas já conhecidas, que podem ser importadas facilmente, foi escolhida para o projeto o modelo de classificação MobileNet, que é uma rede neural que exige menos recursos computacionais em relação a outros modelos. O modelo importado já incluiu pesos treinados com o banco de imagens ImageNet (KERAS, 2019).

Após a criação do modelo MobileNet, últimas camadas são selecionadas para serem treinadas novamente e também é acoplada à saída do modelo mais uma camada cuja quantidade de nós na saída é igual a quantidade de pastas que o usuário do sistema selecionou. Para concluir esta funcionalidade foram utilizados conjuntos de tutoriais disponíveis em sites como Learn OpenCV (GUPTA, 2018) e YouTube (DEEPLIZARD, 2018).

4 USO DA FERRAMENTA

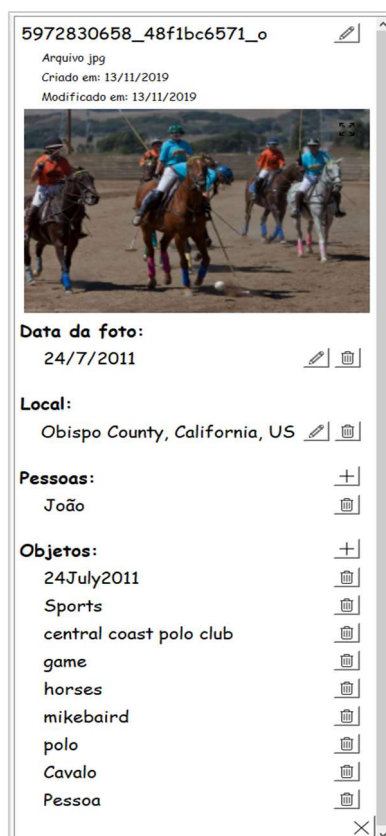
Neste capítulo são apresentadas as funcionalidades principais da aplicação de desktop por meio de recortes de tela obtidos durante a utilização e textos complementares sobre cada função e sua respectiva média de acertos para atividade envolvendo inteligência artificial.

4.1 Metadados

Metadados podem informar mais sobre a imagem, como a data verdadeira que a foto foi tirada, localização, pessoas na foto e palavras chave. O usuário pode editar essas informações disponíveis e pode também acionar as funções de inteligência artificial para adicionar pessoas ou palavras-chave / objetos automaticamente.

Na Figura 12, uma captura de tela do software em funcionamento, é apresentado a funcionalidade de visualização dos metadados existentes em uma imagem, com a possibilidade de edição.

Figura 12 – Visualização de metadados de uma imagem

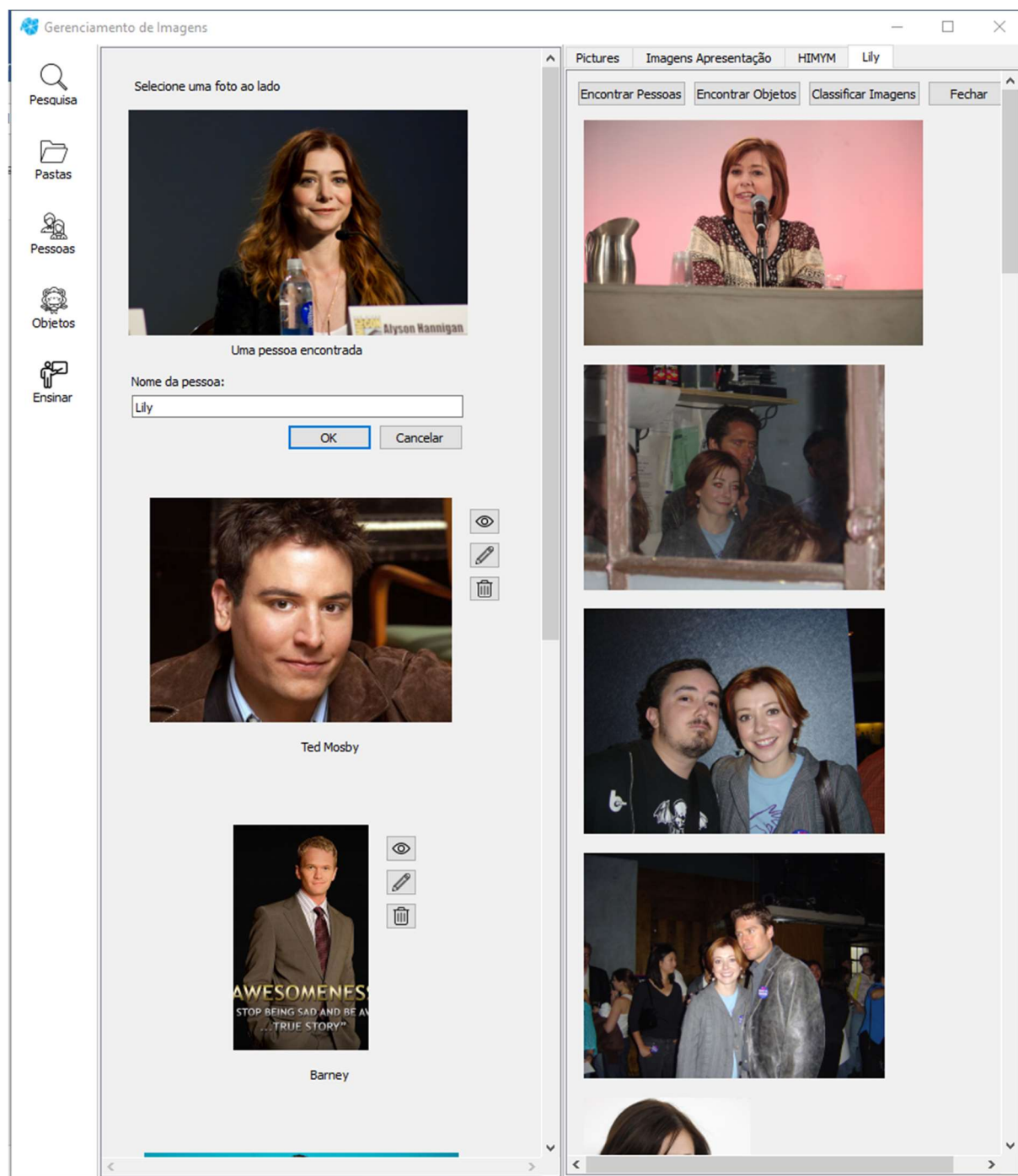


Fonte: Elaborado pelo autor (2019)

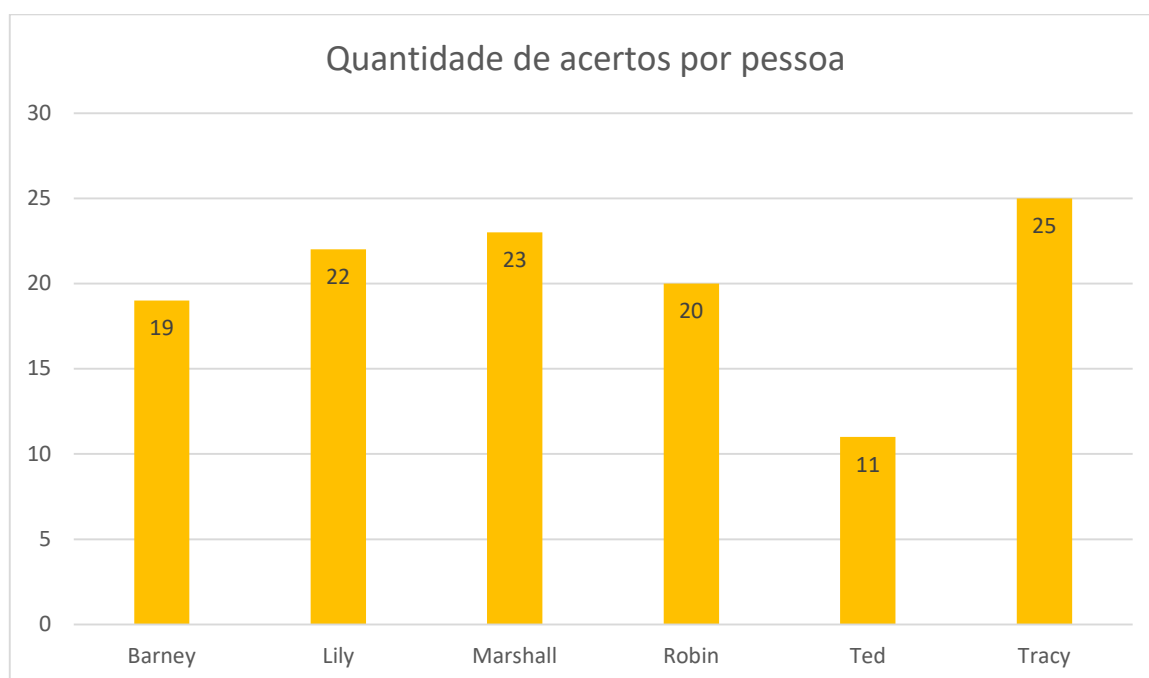
4.2 Pessoas

Com apenas uma foto com o rosto visível de uma pessoa, é possível criar uma assinatura facial e para que o usuário fique ciente das pessoas que o sistema é capaz de procurar, existe uma lista com a foto utilizada e o nome de cada pessoa cadastrada.

Figura 13 – Adicionando uma nova pessoa



Fonte: Elaborado pelo autor (2019)

Figura 14 – Resultado da função de reconhecimento facial

Fonte: Elaborado pelo autor (2019)

De acordo com o resultado acima, numa amostragem contendo 150 imagens o programa foi capaz de inferir corretamente o nome da pessoa que estava na foto em 120 amostras, demonstrando uma acurácia de 80%.

Cada pessoa estava presente em 25 fotos, tiradas de diversos ângulos e de acordo com a observação dos resultados, rostos de pessoas estão de perfil, com o rosto parcialmente coberto, ou muito afastada da câmera tem menos probabilidade de ser serem automaticamente detectados.

Figura 15 – Assertividade da detecção em diferentes ângulos do rosto

Fonte: Elaborado pelo autor (2019)

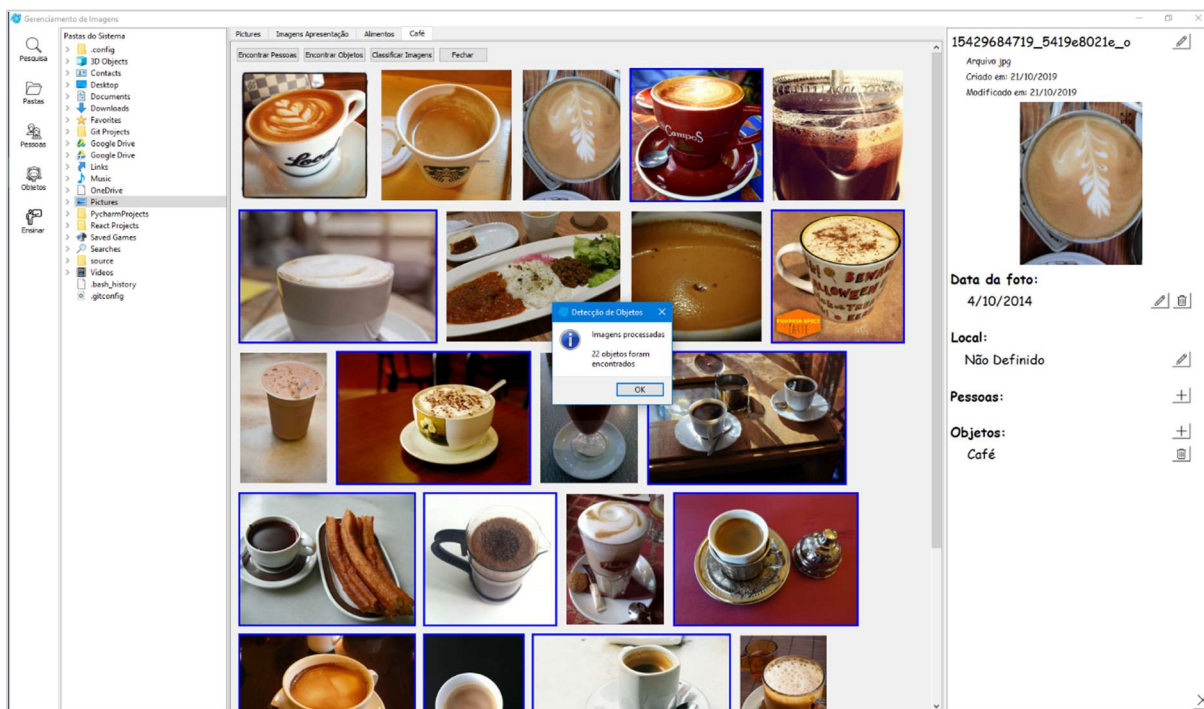
4.3 Objetos

A detecção de objetos é capaz de encontrar mais de 500 tipos de objetos como apresentado na seção 3.3, que versa sobre o modelo YOLOv3.

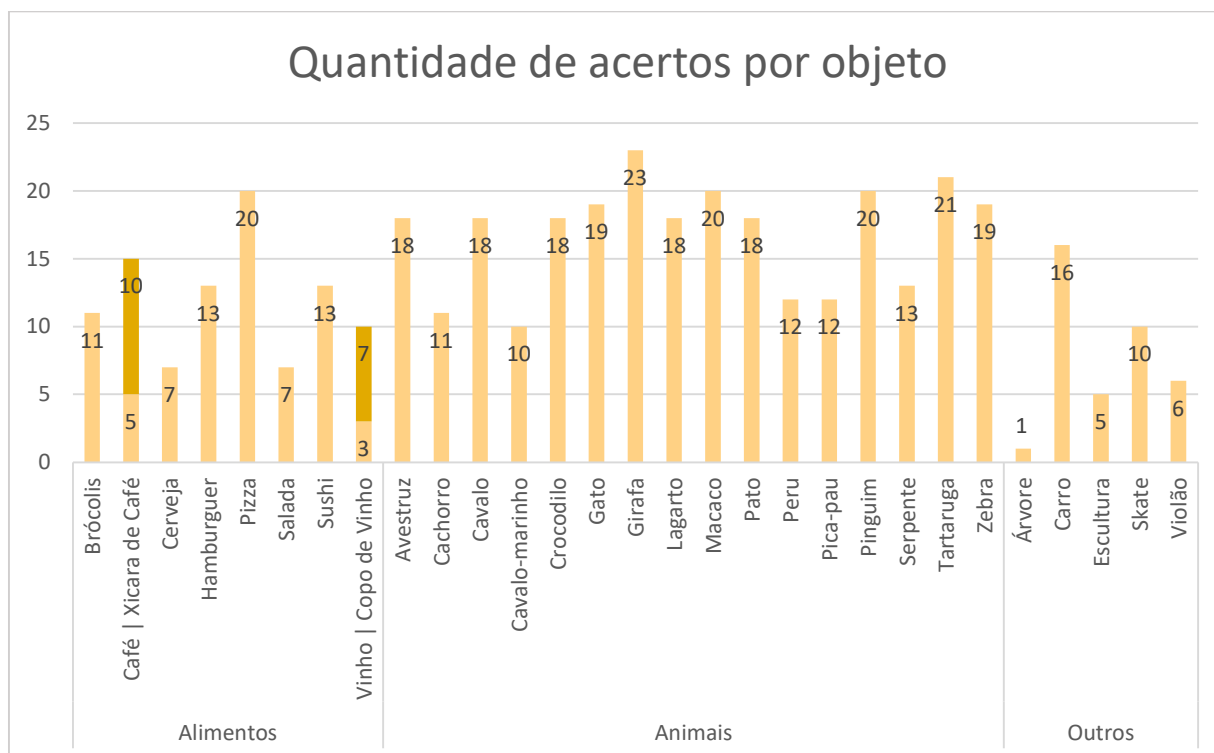
Quando o usuário abre uma imagem pela primeira vez, a função de detecção é chamada e se houver algum objeto conhecido na imagem será adicionado na lista “Objetos”, como é possível observar na parte direita da Figura 16.

Além disso, o usuário pode adiantar a detecção utilizando o botão “Encontrar Objetos” e assim cada imagem presente na pasta aberta será passada para a função de detecção de objetos e as suas respectivas palavras-chave serão adicionadas. A função descrita também está representada na imagem seguinte, com uma mensagem informando o total de objetos encontrados na pasta.

Figura 16 – Encontrando objetos em uma pasta



Fonte: Elaborado pelo autor (2019)

Figura 17 – Teste com objetos diversos

Fonte: Elaborado pelo autor (2018)

No resultado apresentado na Figura 17, a detecção de objetos acertou 55,7% das vezes. Mas vale a pena observar que Xícara de Café e Copo de Vinho são de outra categoria e que na amostragem de validação estava somente Café, ou Vinho. Logo, levando em consideração somente a comparação entre etiqueta da imagem de validação e a palavra-chave inferida, ou seja, sem contar Xícara de Café e Copo de Vinho a acurácia cai para 53,4%.

A amostragem contém principalmente imagens que realmente seriam usadas no cotidiano de uma pessoa para resgatar suas fotos e assim não foi considerado objetos que pouco aparecem em fotos pessoas, apesar de ser possível. Exemplos de objetos que são detectáveis, mas foram desconsiderados na amostragem: Plugues de tomada, equipamentos médico, portas e utensílios de cozinha.

Na competição oficial do banco de imagens utilizado, Open Images, disponível no website Kaggle, como citado na seção 3.3, são contabilizadas todas as 600 classes de objetos e com todas as imagens de validação (diferente do escopo deste projeto onde somente 25 foram selecionadas) e a melhor pontuação foi 58,6%.

Portanto, a acurácia de 53,4% é boa, levando em consideração o proximidade com os resultados dos melhores modelos que existem atualmente para este problema.

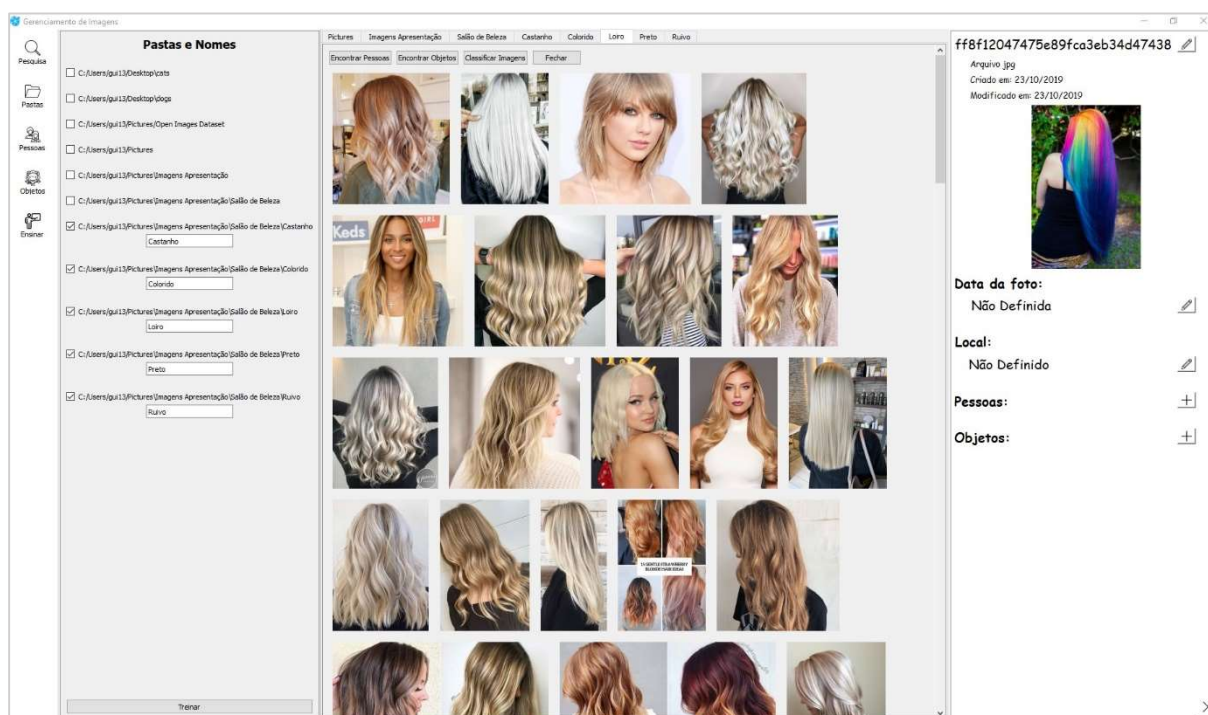
4.4 Classificação

Para o treinamento de um modelo personalizado é preciso que o usuário escolha algumas pastas onde estarão as imagens que serão utilizados para o modelo aprender e diferenciar dos padrões entre pastas.

Quanto mais imagens dentro de cada pasta melhor será o resultado do modelo. Nos exemplos seguintes, com cabelos e posteriormente cenários de um ensaio fotográfico, serão utilizadas 100 imagens para cada classe.

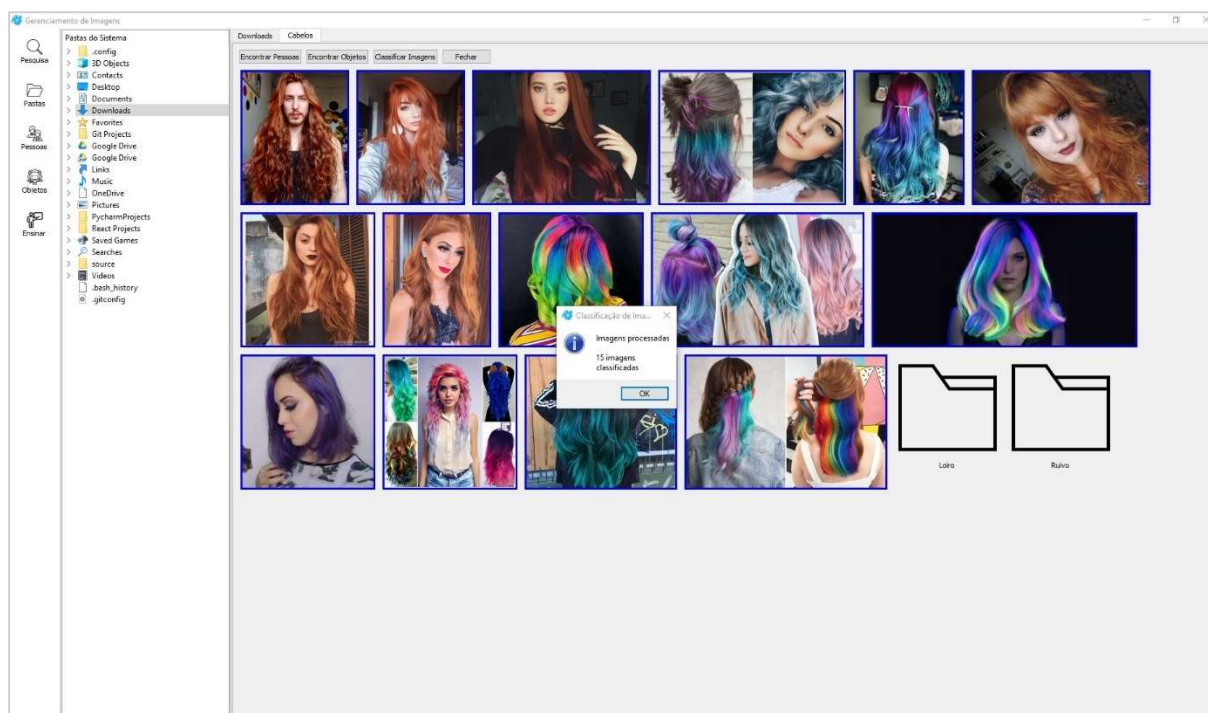
Diferentemente das outras funções evidenciadas nesse projeto, este passo para ensinar o modelo personalizado exige um processamento elevado. Portanto, o programa precisa ficar de 30 minutos e 90 minutos para terminar o processo.

Figura 18 – Seleção das pastas que o modelo utilizará para treino



Fonte: Elaborado pelo autor (2019)

Figura 19 – Processamento de uma pasta com classes aleatórias



Fonte: Elaborado pelo autor (2019)

Durante o desenvolvimento da função de criação de um modelo personalizado de classificação, testes e estudos foram feitos para maximizar o número de acertos da função. O seguinte quadro mostra como se comporta um modelo quando poucas camadas são treinadas (5 camadas):

Quadro 2 – Matriz de confusão com pouco treinamento

		Predição				
		Castanho	Colorido	Loiro	Preto	Ruivo
Real	Castanho	12	0	2	0	6
	Colorido	3	13	0	1	3
	Loiro	6	0	5	0	9
	Preto	9	1	0	5	5
	Ruivo	2	0	0	0	18

Fonte: Elaborado pelo autor (2019)

A acurácia obtida, uma das medidas que podem ser extraídas desta matriz de confusão, cujo cálculo é a razão entre quantidade de acertos pela quantidade total (75), foi de somente 53% (BROWNLEE, 2016; VANETTI, 2007).

O tempo necessário para concluir o treinamento é diretamente proporcional ao número de camadas que serão treinadas do modelo. O número de camadas selecionadas pode aumentar a quantidade de tempo, mas também pode aumentar a acurácia do modelo. Com isso, foram selecionadas 23 camadas do modelo pré-treinado MobileNet que contem 93 camadas.

Esta definição foi mantida como padrão, diminuindo as variáveis para a criação do modelo, que será definida pela quantidade e qualidade de fotos selecionadas pelo usuário. A acurácia obtida com o padrão descrito, e não com os parâmetros de teste, foi de 69,3% e seu resultado é apresentado no Quadro 3.

Quadro 3 – Matriz de confusão: Cores de cabelo

		Predição				
		Castanho	Colorido	Loiro	Preto	Ruivo
Real	Castanho	6	1	1	5	2
	Colorido	1	11	1	1	1
	Loiro	1	1	11	0	2
	Preto	4	0	0	11	0
	Ruivo	1	0	0	1	13

Fonte: Elaborado pelo autor (2019)

Outro exemplo de utilização desta função é apresentado a seguir com tipos e cenário em um ensaio fotográfico. Ensaios fotográficos são na profissão de alguns fotógrafos e estes precisam lidar com esse tipo de foto tanto para armazenamento quanto para fins de portfólio.

Para o teste com ensaios fotográficos, foram selecionadas 75 fotos distribuídas uniformemente em 5 classes, obtendo uma acuraria de 82,7%.

Quadro 4 – Matriz de confusão: fotos de ensaios fotográficos

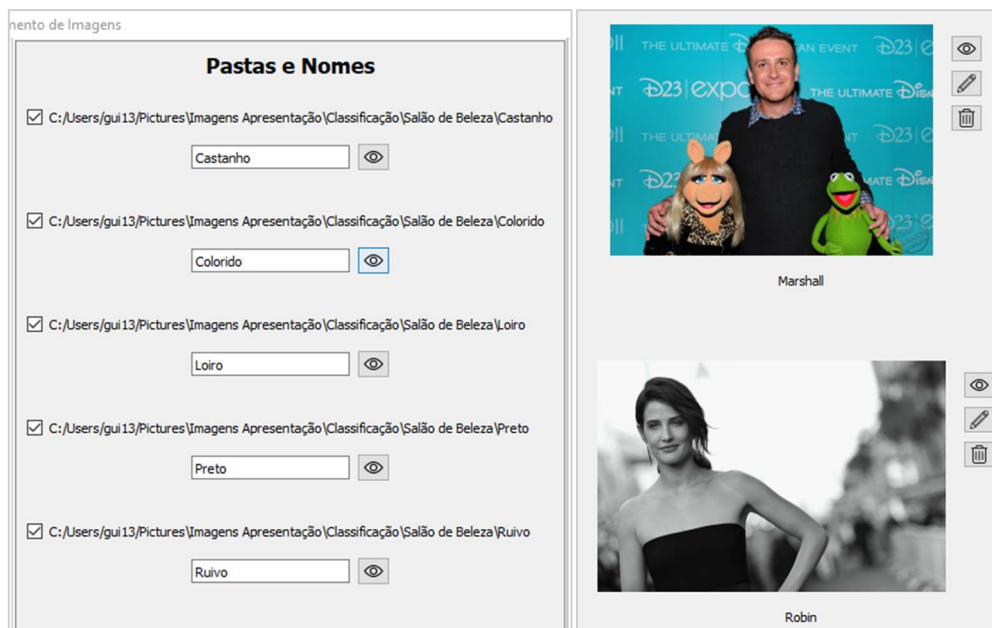
		Predição				
		Bebê	Bosque	Estúdio	Noite	Praia
Real	Bebê	14	1	0	0	0
	Bosque	1	13	0	0	1
	Estúdio	0	0	15	0	0
	Noite	0	6	1	7	1
	Praia	0	2	0	0	13

Fonte: Elaborado pelo autor (2019)

4.5 Resgate de Imagens

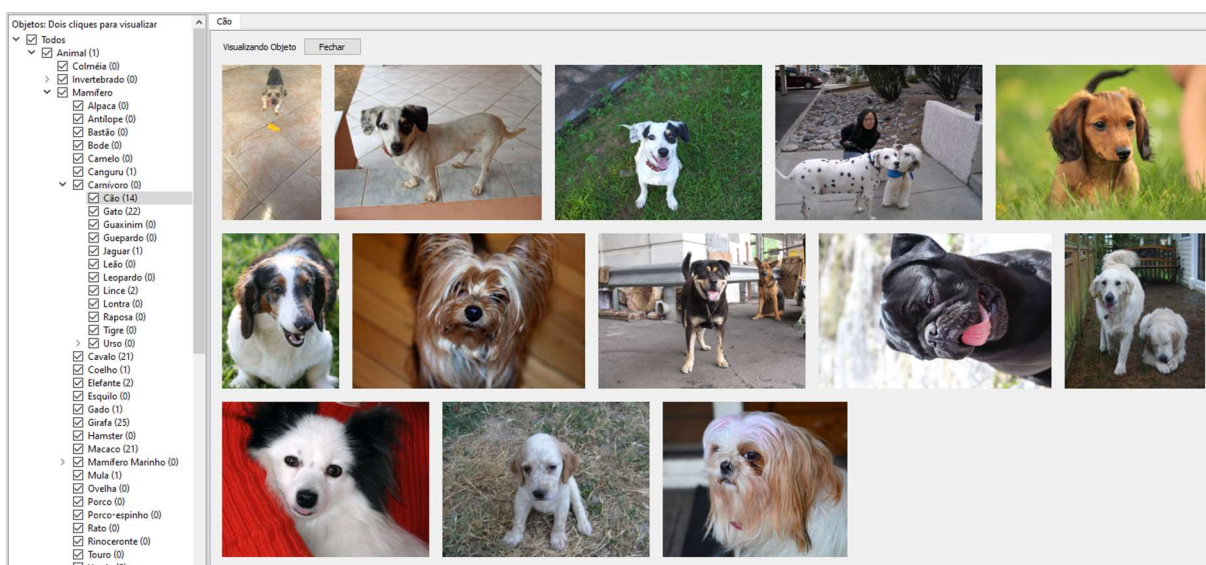
A recuperação das imagens já processadas pelo sistema acontece através da lista de pessoas, de objetos e de classes ensinadas conforme é apresentado na Figura 20 e 21:

Figura 20 – Botão de visualização de imagens em forma de olho



Fonte: Elaborado pelo autor (2019)

Figura 21 – Todas as imagens que possuem determinado objeto



Fonte: Elaborado pelo autor (2019)

5 CONCLUSÃO

O principal aspecto do software proposto é facilitar e automatizar a organização de imagens com o uso de metadados. O software resultante deste trabalho atingiu seus objetivos, passando por desafios e também obtendo aspectos positivos, como será descrito neste capítulo.

Ao decorrer do desenvolvimento deste projeto, conhecimentos foram obtidos e consolidados pelo autor principalmente sobre assuntos relacionados à aprendizado de máquina e suas tecnologias atuais.

O usuário adquire novas experiências ao se deparar com funções pouco utilizadas em seu cotidiano, como controlar um modelo de inteligência artificial e também visualizar fotos com outra perspectiva, levando em consideração as informações agregadas a ela.

Comparando o software desenvolvido com outros softwares, como o Google Photos e o Photos do Windows, todos têm resultados parecidos para encontrar pessoas e objetos, enquanto para a funcionalidade de criar uma classificação própria de imagens por parte do usuário ainda é única neste projeto, não existindo nas outras duas aplicações.

Também é possível observar que o usuário tem mais controle sob as funções existentes, podendo tanto editar metadados livremente quanto corrigir ou escolher o que os modelos de inteligência artificial fazem, diferente de soluções já existentes como Google Photos, conforme foi detalhado na subseção 1.2.2. Tal capacidade se alinha ao objetivo do trabalho, que é fazer com que um usuário final possa utilizar aprendizado de máquina facilmente, sem precisar ser especialista.

A proposta apresentada tem limitações em seu escopo. Entretanto, as possibilidades não abordadas nesse contexto podem ser aplicadas em um projeto futuro, envolvendo mais funcionalidades ou novas plataformas para o software.

5.1 Desafios Encontrados

O principal desafio foi manter o software rápido pois funções que envolvem inteligência exigem um processamento relativamente alto, podendo fazer com que o usuário tenha que esperar alguns segundos para receber uma resposta.

Algumas das bibliotecas e frameworks que foram utilizados exigem parâmetros que melhoram ou pioram a performance da função para determinados casos e como o objetivo do software é trazer a inteligência artificial para um usuário comum, foram mantidas as configurações e parâmetros mais genéricos, para garantir o funcionamento em todos os casos de uso.

Além disso, no treino de classificação de imagens, em que o usuário precisa esperar um tempo, não foi possível comparar o desempenho pois não existe uma solução parecida e o tempo de treinamento depende do hardware do usuário.

5.2 Aspectos Positivos

Como apresentado no capítulo 4, onde é apresentado o uso da ferramenta, os resultados estão alinhados com o objetivo dos modelos pois auxiliam na categorização das imagens com taxas de acertos aceitáveis.

Apesar do treinamento de um modelo personalizado ser um desafio encontrado, continua sendo uma parte necessária antes do uso final da função de classificação. Classificar imagens traz algo não encontrado em outros programas de visualização de imagens, que é a possibilidade de o usuário criar sua própria inteligência artificial. Além das demais funcionalidades permitirem a discussão de uma abordagem interessante, que é a aplicação do aprendizado de máquina em softwares para usuários comuns, que não têm conhecimentos prévios de como criar uma rede neural, algo que ainda exige conhecimentos específicos, como uma linguagem de programação e conceitos fundamentais de aprendizado de máquina.

Neste projeto, tanto o usuário quanto o sistema podem executar a mesma atividade, que é editar metadados, e podem atuar onde exista limitações um do outro. O usuário pode ensinar e corrigir o sistema, e este pode processar imagens de forma mais rápida que o usuário.

5.3 Melhorias Futuras

Apesar de softwares e frameworks serem mais robustos em versões para desktop, também há a possibilidade de transferir o software para uma versão mobile.

Também há a possibilidade de criação de funcionalidades novas para resgates específicos de imagens, como por local e data, apesar de já ser possível

em soluções existentes pesquisar as imagens com metadados editados do software deste trabalho, já que são compatíveis.

Também seria possível pensar na possibilidade de aprendizado não-supervisionado, em que não é preciso “ensinar” o sistema. Isso pode facilitar o uso da ferramenta ao decidir automaticamente o que categorizar com alguns algoritmos de agrupamento (Daumé III, 2015, p. 166).

REFERÊNCIAS BIBLIOGRÁFICAS

- ACDSee. **Reconhecimento facial do software ACDSee**. 9 de jan. de 2019. Disponível em: <<https://www.youtube.com/watch?v=ixVyEjPyNs0&t=57s>>. Acesso em: 20 de nov. de 2019.
- ADOBE ACROBAT. **Propriedades e metadados do PDF**. 2017. Disponível em: <<https://helpx.adobe.com/br/acrobat/using/pdf-properties-metadata.html>>. Acesso em: 19 de nov. de 2019.
- BROWNLEE, Jason. **What is a Confusion Matrix in Machine Learning**. 2016. Disponível em: <<https://machinelearningmastery.com/confusion-matrix-machine-learning/>>. Acesso em: 19 de nov. de 2019.
- CAMERA & IMAGING PRODUCTS ASSOCIATION. **Exchangeable image file format for digital still cameras: Exif Version 2.3**. abril de 2010. Disponível em: <http://www.cipa.jp/std/documents/e/DC-008-2012_E.pdf>. Acesso em: 19 de nov. de 2019.
- CYBER KIBBUTZ. Imagens que envolvem o mesmo domínio: Números.
- DAUMÉ III, Hal. **A Course in Machine Learning**. 1. ed. 2015.
- DEEPLIZARD. **Sign language image classification - Fine-tuning MobileNet with Keras**. 9 de jun. de 2018. Disponível em: <<https://youtu.be/F38J9R5wuqo>>. Acesso em: 14 de nov. de 2019.
- FARHADI, Ali; REDMON, Joseph. **YOLO9000: Better, Faster, Stronger**. 25 de dez. de 2016. Disponível em: <<https://arxiv.org/pdf/1612.08242.pdf>>. Acesso em: 19 de nov. de 2019.
- GEITGEY, Adam. **face_recognition 1.2.3. Recognize faces from Python or from the command line**. 21 de ago. de 2018. Disponível em: <https://pypi.org/project/face_recognition/>. Acesso em: 14 de nov. de 2019.
- GOOGLE PHOTOS. **Google Fotos: Armazenamento gratuito e organização automática para todas as suas memórias**. 2019. Disponível em: <<https://www.google.com/photos/about/>>. Acesso em: 18 de nov. de 2019.
- GOOGLE RESEARCH. **Google AI Open Images - Object Detection Track**. 2018. Disponível em: <<https://www.kaggle.com/c/google-ai-open-images-object-detection-track/overview>>. Acesso em: 18 de nov. de 2019.
- GOOGLE WEBMASTERS. **Does Google use EXIF data from pictures as a ranking factor?** 24 de fev. de 2014. Disponível em:

<<https://www.youtube.com/watch?v=GMf6FmRus2M>>. Acesso em: 18 de nov. de 2019.

GUPTA, Vikas. **Keras Tutorial: Transfer Learning using pre-trained models.**

Learn OpenCV, 3 de jan. de 2018. Disponível em:

<<https://www.learnopencv.com/keras-tutorial-transfer-learning-using-pre-trained-models>>. Acesso em: 14 de nov. de 2019.

IPTC. **IPTC Photo Metadata Standard.** 2017a. Disponível em:

<<https://iptc.org/standards/photo-metadata/iptc-standard/>>. Acesso em: 19 de nov. de 2019.

IPTC. **IPTC: mapping of metadata between the IPTC Photo Metadata Standard 2017.1 and NewsML-G2 2.25.** 2017b. Disponível em:

<<https://iptc.org/std/photometadata/specification/mapping/iptc-pmd-newsmlg2.html>>. Acesso em: 19 de nov. de 2019.

KERAS Documentation. **Keras: The Python Deep Learning library.** 2019.

Disponível em: <<https://github.com/keras-team/keras/>>. Acesso em: 14 de nov. de 2019.

KHOURY, Rita El. **Google Photos rolls out manual face tagging: Tutorial, benefits, and a big asterisk.** 27 de nov. de 2019. Disponível em:

<<https://www.androidpolice.com/2019/11/27/google-photos-rolls-out-manual-face-tagging-tutorial-benefits-and-a-big-asterisk/>>. Acesso em: 3 de dez. de 2019.

KUZNETSOVA, Alina *et al.* **The Open Images Dataset V4. Unified image classification, object detection, and visual relationship detection at scale.** 2 de nov. de 2018. Disponível em <<https://arxiv.org/pdf/1811.00982.pdf>>. Acesso em: 19 de nov. de 2019.

MOREY, Raelene. **How Image Metadata Can Impact Web Performance and Security.** 24 de jul. de 2018. Disponível em:

<<https://wp-rocket.me/blog/image-metadata-can-impact-web-performance-security/>>. Acesso em: 19 de nov. de 2019.

NELL, Waldo. **Exemplo de informações EXIF.** 22. De fev. de 2018. Disponível em:

<<https://www.dpreview.com/forums/thread/4257730>>. Acesso em: 20 de nov. de 2019.

NORCROSS, Amanda. **Uma pessoa, um cachorro e um carro na foto.** 30 de set. de 2015. Disponível em:

<<https://coastguard.dodlive.mil/2016/03/ruff-duty-coast-guard-canine-explosive-detection-teams/>>. Acesso em: 20 de nov. de 2019.

NORTON. **How does facial recognition works?** 2019. Disponível em: <<https://us.norton.com/internetsecurity-iot-how-facial-recognition-software-works.html>>. Acesso em: 19 de nov. de 2019.

RAHUL. **Aplicativo Photos do Windows**. Disponível em: <<https://mspoweruser.com/microsoft-photos-app-updated-with-new-graphic-layout-and-more/>>. Acesso em: 20 de nov. de 2019.

RUDER, Sebastian. **Imagens que envolvem o mesmo domínio: Números**. 21 de mar. De 2017. Disponível em: <<https://ruder.io/transfer-learning/index.html>>. Acesso em: 20 de nov. de 2019.

SZELISKI, Richard. **Computer Vision: Algorithms and Applications**. 2010, p. 685. Springer. Disponível em: <http://szeliski.org/Book/drafts/SzeliskiBook_20100903_draft.pdf>. Acesso em: 19 de nov. de 2019.

VANETTI, Marco. **Confusion matrix online calculator**. 2007. Disponível em: <<http://www.marcovanetti.com/pages/cfmatrix/?noc=5>>. Acesso em: 19 de nov. de 2019.

ZEMAN, Jan. **Understanding EXIF: What Metadata Is and How to Use It**. 6 de nov. de 2018. Disponível em: <<https://learn.zoner.com/understanding-exif-what-metadata-is-and-how-to-use-it/>>. Acesso em: 20 de nov. de 2019.