
## Instituto Universitário de Lisboa

Departamento de Ciências e Tecnologias da Informação


# Identificação Automática de Plantas Invasoras em Imagens Aéreas


Carolina do Carmo Lages Gonçalves


Dissertação submetida como requisito parcial para obtenção do grau de

**Mestre em Engenharia Informática**


**Orientador:**

Professor Doutor Pedro Figueiredo Santana

ISCTE-IUL

**Co-orientador:**

Professor Doutor Tomás Gomes da Silva Serpa Brandão

ISCTE-IUL


Setembro, 2019

# *Resumo*

As espécies invasoras são conhecidas pela sua rápida disseminação, originando perda de biodiversidade das zonas invadidas, tornando-se importante realizar a monitorização das áreas florestais para o controlo destas espécies. Esta dissertação apresenta uma arquitetura para a deteção da espécie invasora *Acacia Longifolia* em imagens aéreas, particularmente relevante em Portugal. A arquitetura inclui a captura de imagens aéreas através de veículos aéreos não tripulados (VANTs), o pré-processamento das imagens e a divisão do conjunto de dados em treino, validação e teste. Também inclui uma rede neuronal convolucional (RNC) treinada para a classificação automática desta espécie invasora, nas imagens aéreas adquiridas. Testaram-se duas configurações da RNC, cuja arquitetura difere na última camada responsável pela classificação de amostras com 100 x 100 píxeis, obtidas por uma janela deslizante ao longo das imagens capturadas. Uma das redes classifica segundo nove classes (e.g., Acacia L., Vegetação, Estrada), sendo que a classificação obtida é convertida numa classificação binária através da sua matriz de confusão, tendo apresentado uma taxa de acerto de 98.5% utilizando o conjunto de teste. O segundo modelo foi treinado para a classificação binária relativa à presença de Acácia L., alcançando-se um desempenho de 98.7%. Os resultados mostram que a multi-classificação não prejudica o desempenho na deteção da *Acacia Longifolia* e fornece ao VANT informação adicional relativa ao ambiente. Por último, desenvolveu-se uma abordagem para melhorar a taxa de acerto, recorrendo a um especialista para verificar as previsões do sistema, ponderando-se o benefício em melhorar o desempenho com o custo de chamar o especialista.

**Palavras-chave:** Aprendizagem Automática Profunda, Redes Neuronais Convolucionais, Plantas Invasoras, *Acacia Longifolia*

# *Abstract*

Invasive species are known for their rapid dissemination, involving the loss of biodiversity in affected areas, becoming important to monitor the forest areas in order to control these species. This dissertation presents an architecture for the detection of the invasive species *Acacia Longifolia* in aerial images, particularly relevant in Portugal. The architecture includes capturing aerial images through unmanned aerial vehicles (UAVs), preprocessing the images and splitting the data into training, validation and testing sets. It also includes a trained convolutional neuronal network for automatic species classification based on the acquired aerial images. Two models were built, whose architecture differs in the last layer responsible for classifying samples with 100 x 100 pixels, obtained by a sliding window along the high-resolution images. One of the networks classifies according to nine classes (e.g., Acacia L., Vegetation, Roadway), and the obtained classification is then converted into a binary classification through the confusion matrix, having an accuracy of 98.5% for testing set. The second model was trained for binary classification for the presence of Acacia L., achieving an accuracy of 98.7% for the test set. The results show that the multi-classification does not hamper the performance of *Acacia Longifolia* detection and provides UAV with additional environmental information. Finally, an approach has been developed to improve the accuracy of the system by calling an expert to review the predictions produced by the system, balancing the expected benefit of accuracy improvement with the cost of calling the expert.

**Keywords:** Deep learning, Convolutional Neural Network, Invasive species, *Acacia Longifolia*

# *Agradecimentos*

Gostaria de agradecer aos meus pais pela confiança e apoio fornecido. Agradeço ao meu orientador Professor Pedro Santana por toda a ajuda e esclarecimento dado ao longo do ano e, também, ao meu co-orientador Professor Tomás Brandão pelas suas opiniões construtivas. Ao meu namorado e colega de trabalho, Bruno Teles, que me acompanhou em todo este processo e me foi ajudando quando precisava, acreditando sempre que iria conseguir terminar. Agradeço à empresa IntRoSys S.A. por me ter fornecido o conjunto de dados, essencial para conseguir realizar a dissertação, e pelo auxílio prestado através da disponibilização de recursos computacionais (GPU) que permitiram reduzir o tempo de computação dos resultados. Por fim, a dissertação foi realizada nas instalações do Instituto de Telecomunicações no ISCTE-IUL.

# Conteúdo

# Lista de Figuras

# Lista de Tabelas

# Acrónimos

**VANTS** Veículos Aéreos Não Tripulados

**SVM** Support Vector Machine

**RNA** Redes Neuronais Artificiais

**RNC** Redes Neuronais Convolucionais

**RCC** Rede Completamente Convolucional

**HOG** Histogram of Oriented Gradients

**RGB** Red-Green-Blue

**VPN** Vehicle Proposal Networks

**MS-OPN** Multi-Scale Object Proposal Network

**AODN** Accurate Object Detection Network

**SSMD** Single Shot MultiBox Detector

**SSH** Single Stage Headless

# Capítulo 1

# Introdução

As espécies invasoras são conhecidas pelo seu rápido crescimento, alterando e reduzindo a biodiversidade da zona invadida. Estas características não só acarretam impactos negativos nos ecossistemas, como também prejuízos económicos, nomeadamente, quando invadem campos agrícolas e áreas florestais, competindo pela aquisição de recursos e implicando perda das espécies existentes [Marchante et al., 2014].

A espécie *Acacia Longifolia* é uma árvore de pequeno porte, considerada como uma espécie invasora em Portugal e noutros países da Europa, devido à produção em abundância de sementes que geram um aumento na proliferação destas espécies, ocupando e alterando o habitat natural das espécies nativas [Invasoras.pt, 2019]. É reconhecida devido às características das suas flores em forma de espigas amarelas. As técnicas usadas para o controlo desta espécie são dispendiosas e por vezes não eliminam totalmente o problema. Uma das técnicas existentes consiste na aplicação de um agente de controlo biológico nas áreas afetadas, que impede o crescimento das sementes ou que se alimenta das mesmas prevenindo a germinação de novas plantas desta espécie, sendo essencial monitorizar as áreas durante e após a intervenção para avaliar a evolução das espécies.

A deteção remota usando imagens de satélites é uma das técnicas clássicas usadas para mapear a vegetação e controlar a ocupação das áreas que, combinando com

métodos de aprendizagem automática, como Support Vector Machine (SVM) e Redes Neuronais Artificiais (RNA) possibilita, por exemplo, classificar as várias espécies da flora automaticamente [Martins, 2012]. No entanto, os métodos de aprendizagem automática obrigam a extração prévia e manual das características para serem aplicadas nesses classificadores e, por vezes, de forma a melhorar o desempenho do modelo, é necessário aplicar informações adicionais, como demonstrado na investigação de Mendes e Dal Poz [Mendes and Dal Poz, 2011]. De forma a evitar este pré-processamento da seleção das características recorre-se à aprendizagem automática profunda [1] que é uma subárea da aprendizagem automática. Na aprendizagem profunda utilizam-se redes neuronais complexas para a extração automática de características e para a classificação. Neste caso, como envolve processamento e análise de imagens, recorre-se a um modelo de aprendizagem automática profunda designado por redes neuronais convolucionais (RNC) para realizar a deteção e reconhecimento de objetos de classes diferentes. As RNC permitem realizar a classificação de imagens em diferentes áreas, tais como a ocupação do território [Castelluccio et al., 2015], a deteção de automóveis [Audebert et al., 2017], a contagem de árvores ou de animais [Liu et al., 2018] e a classificação de plantas agrícolas [Huang et al., 2018].

Com o desenvolvimento e evolução da tecnologia surgiram os veículos aéreos não tripulados (VANTs) que capturam imagens aéreas e permitem, também, construir mapas de vegetação, fornecendo informações sobre a distribuição de espécies numa dada área [Kaneko and Nohara, 2014] e [Cruzan et al., 2016]. Esta última tecnologia apresenta vantagens em relação ao uso de satélites, como tendo um menor custo de aquisição e de utilização, para além de permitir uma obtenção rápida de imagens da área em estudo [Cruzan et al., 2016]. Esta união de tecnologias permite estudar a ocupação do território [Ham et al., 2018] e mapear e detetar espécies da flora [Wang et al., 2019]. No entanto, a deteção e classificação de espécies invasoras da flora em imagens aéreas capturadas por VANTs, é uma área com uma exploração inexistente. Por estas razões, torna-se importante desenvolver ferramentas que promovam e facilitem a prevenção, controlo e gestão sustentável das espécies

[1] Do Inglês *Deep Learning*

invasoras, permitindo ações de resposta imediatas ao problema e recuperando os habitats afetados, recorrendo às tecnologias referidas anteriormente.

Nesta dissertação foi desenvolvida uma arquitetura que permite o reconhecimento da espécie invasora *Acacia Longifolia* através do desenvolvimento de duas redes neuronais convolucionais para o reconhecimento da espécie invasora em estudo. Uma das redes foi treinada para realizar classificação binária entre as classes Acacia L. e Não Acacia L. e a outra rede, para a classificação segundo nove classes (Acacia L., Pinheiro, Estrada, Outras amarelas, Sobreiro, Outros componentes, Restos de árvores, Pequenas ervas e Vegetação) cujo resultado foi convertido para binário, para perceber se o treino de multi-classes dificultaria o reconhecimento da espécie invasora. Através dos resultados dos dois modelos desenvolvidos validou-se o elevado desempenho da aplicação de RNC para a deteção da espécie invasora *Acacia Longifolia* em imagens aéreas capturadas por VANTs e concluiu-se que o treino para a distinção de nove classes não prejudica o desempenho do modelo na tarefa principal. Desta forma, foi colmatada a necessidade de desenvolver uma ferramenta para a deteção precoce e automática das espécies invasoras.

Para melhorar a taxa de acerto da RNC pode-se reajustar os hiperparâmetros do modelo, sendo, no entanto, necessário re-treinar a rede. Este processo, para além de ser lento, pode induzir a rede a ajustar-se em excesso aos dados, perdendo a capacidade de generalizar corretamente quando exposta a dados não observados durante a fase de treino. Para se conseguir melhorar a classificação obtida nos novos dados foi desenvolvido um mecanismo que recorre a um especialista para verificar algumas das classificações previstas pela rede. O algoritmo proposto permite escolher as previsões que devem ser classificadas pelo especialista tendo em conta o compromisso entre a vantagem de melhorar a classificação prevista e o custo da chamada a um especialista para a verificação das previsões. Esta abordagem permite melhorar a classificação tendo em conta o fim da aplicação do modelo de classificação definido pelo utilizador.

## 1.1 Questões de Investigação

Com base no problema referido anteriormente colocam-se as seguintes perguntas de investigação:

- A aplicação de redes neuronais convolucionais para a deteção de plantas invasoras em imagens aéreas, contribui para o aumento de desempenho comparativamente às técnicas existentes?
- Qual a topologia da rede neuronal convolucional que melhor se ajusta ao problema em questão?
- Que tipo de métodos de regularização permitem tornar o modelo de rede neuronal convolucional mais preciso, tendo em conta o problema?
- A utilização de múltiplas classes influencia a taxa de reconhecimento correto da espécie *Acacia Longifolia*?
- Será possível melhorar, de forma eficiente, a taxa de acerto da rede após treinada, recorrendo à ajuda de um especialista?

## 1.2 Objetivos

Esta dissertação visa o desenvolvimento de uma ferramenta para detetar e segmentar espécies de plantas invasoras em imagens aéreas capturadas por um veículo aéreo não tripulado (VANT) através de técnicas de aprendizagem profunda, em particular, as redes neuronais convolucionais (RNC). A espécie em estudo é a *Acacia Longifolia* que é uma das espécies invasoras que apresenta maior prejuízo nos ecossistemas em Portugal. Os objetivos do trabalho são os seguintes:

- Modelar um classificador usando os pacotes de software livre Keras, Tensorflow e OpenCV;
- Treinar e testar um classificador para o reconhecimento da espécie invasora *Acacia Longifolia* usando um conjunto de dados recolhidos por um VANT.

## 1.3 Estrutura do documento

Este documento está dividido por mais cinco capítulos: no **Capítulo 2** está presente uma breve informação teórica relevante sobre as redes neuronais convolucionais. No **Capítulo 3** são referidas as investigações relacionadas com o reconhecimento e deteção automática de objetos em imagens aéreas, enfatizando a aplicação de RNC na área da flora. No **Capítulo 4** é descrita a abordagem proposta para a resolução do problema, nomeadamente a aquisição, composição e divisão do conjunto de imagens, as arquiteturas das redes, configuração do treino e um estudo empírico para aumentar a taxa de acerto. No **Capítulo 5** são apresentados os resultados obtidos, tal como o desempenho dos modelos no reconhecimento da espécie invasora *Acacia Longifolia*, o resultado da segmentação de imagens de alta resolução e os resultados do algoritmo que permite melhorar a taxa de acerto recorrendo a um especialista que verifica as previsões filtradas por um limiar de confiança calculado pelo algoritmo. Por último, são apresentadas as considerações finais e sugestões para trabalho futuro (**Capítulo 6**).

# Capítulo 2

# Fundamentos da Aprendizagem Profunda

A aprendizagem profunda é uma subcategoria da aprendizagem automática relacionada com o desenvolvimento de redes neuronais com várias camadas. Os algoritmos de aprendizagem automática conseguem aprender a classificar diferentes tipos de informação, como som, texto e imagens. O processamento de imagens envolve a utilização de redes neuronais convolucionais que são um tipo de redes neuronais artificiais (RNA) do ramo da aprendizagem automática. A redes neuronais convolucionais permitem a extração automática de características [1], que representam a informação recebida por estes modelos.

A redes neuronais foram desenvolvidas com base no funcionamento do cérebro [Mendes and Dal Poz, 2011]. Resumidamente, os neurónios (Figura 2.1a) são os principais componentes na transmissão de sinais no cérebro e são compostos por: dendrites, que recebem sinais de outros neurónios; um corpo celular que processa esses sinais e determina se transmite ou não o sinal aos neurónios seguintes desencadeando um impulso elétrico; e uma arborização terminal por onde se transmite o sinal aos neurónios seguintes através de sinapses. Cada neurónio recebe sinais de vários neurónios e transmite para mais que um neurónio, originando uma rede de neurónios.

[1] Do Inglês *features*.

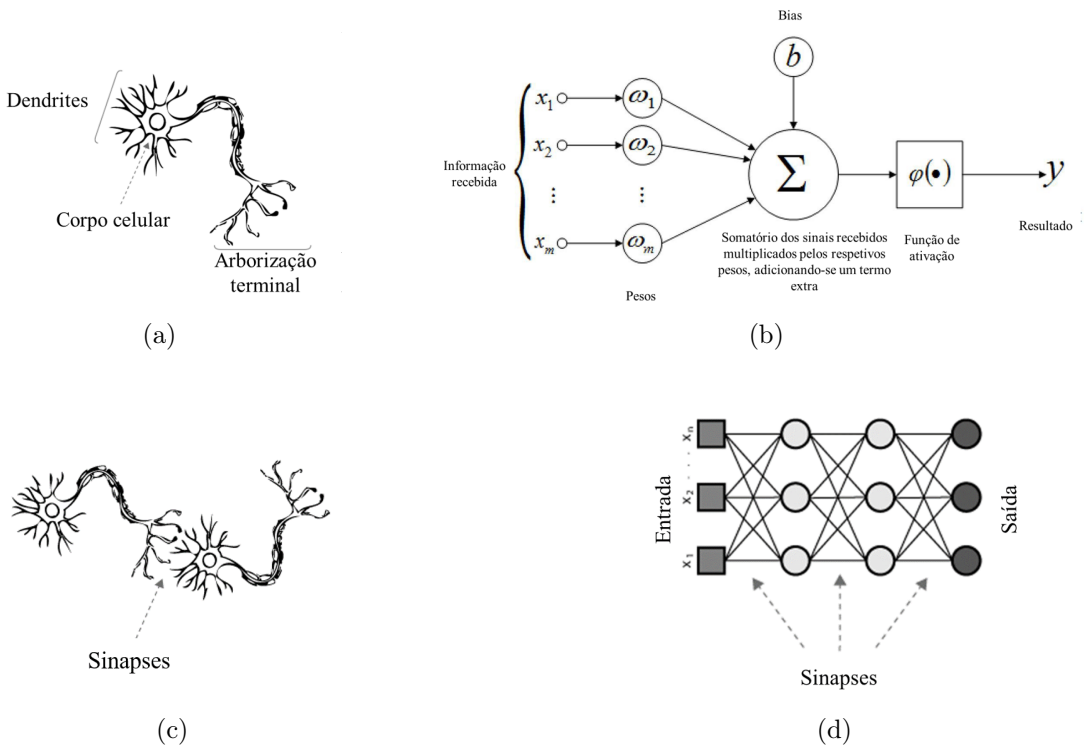



FIGURA 2.1: Paralelismo entre neurónios e redes biológicas e artificiais. (a) representa a estrutura de um neurónio biológico e (b) a de um neurónio artificial. As figuras (c) e (d) ilustram as redes compostas por neurónios biológicos e artificiais, respetivamente. Figura adaptada da Figura 2 do artigo de Wen [Wen and Geomatics, 2016].

As RNA são compostas por unidades interligadas designadas por neurónios (Figura 2.1a) que recebem informações de várias entradas representadas por $x$ nas Figuras 2.1b e 2.1d, associadas a um peso, $w$ na Figura 2.1b, que representam o valor da informação recebida para a classificação. O neurónio realiza o somatório desses pesos multiplicados pela respetiva informação e adiciona um termo extra designado por *bias* permitindo ao modelo uma maior flexibilidade para aprender os dados. O resultado do somatório é recebido por uma função de ativação que coordena o valor de saída de cada unidade tendo em conta o valor do termo extra anterior que permite variar o valor limite para ativar ou não o neurónio. Os valores resultantes do somatório sofrem uma transformação não linear através das funções de ativação para garantir que a rede, durante a aprendizagem, não seja sensível a pequenas variações, para além de que, sem a não-linearidade da função de ativação não seria possível aprender padrões não-lineares. Tal como no funcionamento do cérebro, as RNA são compostas por vários neurónios ligados entre si

formando uma rede composta por camadas de neurónios. A primeira camada recebe a informação e designa-se por camada de entrada, a camada de saída retorna a classificação prevista, ou seja, um conjunto de probabilidades associadas a cada classe que representam determinados objetos. As restantes camadas designam-se por camadas escondidas e contribuem também para o cálculo da classe a ser prevista.

As RNC [LeCun et al., 1999] (Figura 2.2) diferem das redes neuronais clássicas, no número de ligações entre unidades. Nas RNC, cada unidade recebe uma fração do resultado da saída da unidade anterior, determinada pela dimensão do filtro que é uma matriz de pesos para a deteção de característica e é definido em cada camada [Branco, 2017]. Por outro lado, nas redes neuronais artificiais não convolucionais os neurónios de cada camada interligam-se com todos os neurónios da camada seguinte [Castelluccio et al., 2015]. Deste modo e comparando uma RNA com os neurónios todos interligados e com o mesmo número de camadas que uma RNC, a utilização de redes neuronais não convolucionais na classificação de imagens implica num elevado custo computacional, tornando o treino da rede impraticável, pois o número de parâmetros a serem aprendidos é superior do que nas RNC.

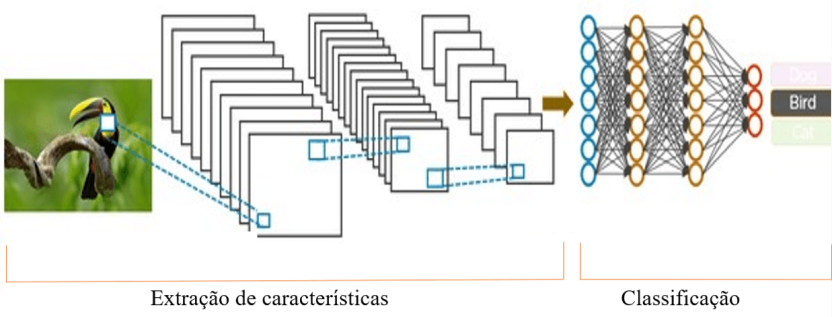



FIGURA 2.2: Topologia de uma rede neuronal convolucional adaptada da imagem pertencente à página web [LABS, 2019]. A deteção de características é realizada pelas camadas de convolução seguidas da função de ativação e das camadas de sub-amostragem. As informações obtidas pelas camadas referidas anteriormente são recebidas pelas camadas completamente conetadas para a classificação das imagens.

As redes neuronais convolucionais são compostas por camadas convolucionais responsáveis pelas operações de convolução entre a imagem recebida, que é uma matriz composta por três matrizes com informação RGB separadas, e os filtros

de convolução para a descoberta de padrões da imagem. Estes padrões tornam-se mais específicos ao longo das camadas da rede. Na convolução é computado o produto interno na imagem recebida pela rede através de uma janela deslizante com dimensão e os pesos do filtro de convolução, como ilustra a Figura 2.3.

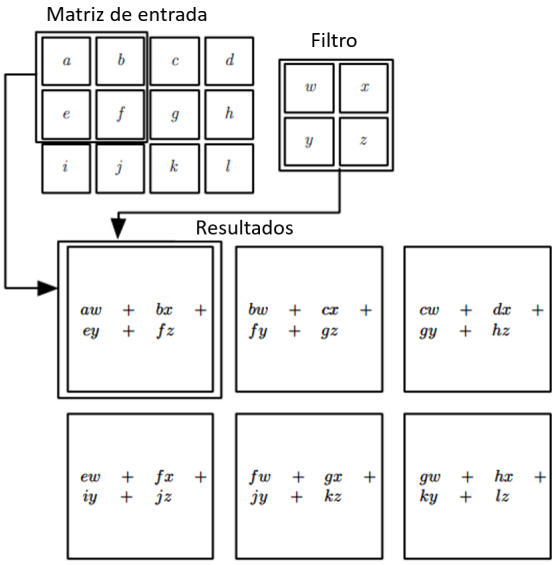



FIGURA 2.3: Operação de convolução entre a matriz recebida e o filtro de convolução. Figura retirada do livro [Goodfellow et al., 2016].

A dimensão resultante será inferior à da imagem original, perdendo-se alguma informação presente nos limites da imagem. Para corrigir estes problemas, adiciona-se píxeis com valor zero [2] à volta dos limites da imagem original. São adicionados os píxeis necessários para se preservar a dimensão após a convolução.

As RNC também são compostas por camadas de sub-amostragem com o objetivo de reduzir o número de píxeis da matriz resultante da convolução e consequentemente, o número de parâmetros a serem aprendidos. Esta sub-amostragem pode ser realizada através da seleção do valor máximo entre o conjunto de píxeis abrangidos pela dimensão do filtro de sub-amostragem [3] ou então, realiza-se a média dos pesos abrangidos pelo filtro [4]. Esta sub-amostragem e a operação de convolução são realizadas ao longo de toda a matriz recebida, sendo que a rapidez do seu progresso é influenciada pelo número de píxeis que o filtro transita pela imagem [5],

[2] Designado por *zero-padding*
[3] Designada por *max pooling*
[4] Designada por *average pooling*
[5] Designado por *stride*

após cada operação, sendo que quanto maior for o seu valor menor será o tempo de computação. Por exemplo, na Figura 2.4, a matriz recebida apresenta uma dimensão de 4 x 4 píxeis e a sub-amostragem é realizada por um filtro 2 x 2 com deslocamento de dois, ou seja, cada sub-amostragem é feita entre quatro píxeis escolhendo-se o máximo do conjunto. Após cada operação, o filtro desloca-se dois píxeis na horizontal até terminar a linha e verticalmente pelas colunas.

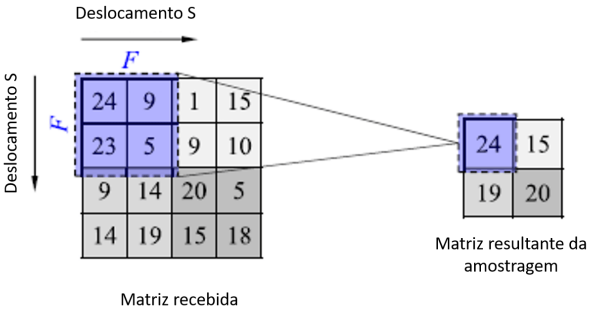



FIGURA 2.4: Operação de sub-amostragem resultante no máximo entre a matriz recebida e o filtro 2x2 com deslocamento S=2

Após a camada de convolução e de sub-amostragem, os valores resultantes sofrem uma transformação não linear através das funções de ativação [Goodfellow et al., 2016] para garantir que a rede, durante a aprendizagem, não seja sensível a pequenas variações, para além de que, sem a não-linearidade da função de ativação não seria possível aprender padrões não-lineares . Existem vários tipos de funções de ativação com as suas vantagens e desvantagens. O mais comum em aprendizagem profunda é do tipo **ReLU** [Patterson and Gibson, 2017], utilizado principalmente, nas camadas intermédias e converte os pesos negativos para zero, sendo que os positivos se mantêm inalterados. As unidades com peso nulo não serão consideradas pelas próximas camadas, para a classificação, decrementando o tempo de treino. Por fim, nas camadas completamente conectadas começa-se por juntar as informações recebidas num único vetor unidimensional [Bergado et al., 2016] para a posterior classificação através da função de ativação *softmax*. *Softmax* é utilizado em previsões de duas ou mais classes, retornando um vetor com as probabilidades do objeto pertencer a cada classe.

**Aprendizagem das RNC:** Uma rede neuronal convolucional recebe na entrada uma imagem, que será representada por matrizes com os valores RGB ou outras

dimensões de cores, separadas. Essa informação atravessa a rede, sofrendo operações como convolução, sub-amostragem e transformações não lineares, até alcançar as últimas camadas para a classificação e onde se computa a função de perda, $L$, que informa a distância entre a classe prevista pela rede e a verdadeira de acordo com:

$$L(y', y) = -\sum_{i=1}^{n} y_i log(y'_i), \tag{2.1}$$

sendo $y'$ a classe prevista, $y$ a classe original e $n$ corresponde ao número total de amostras de treino. Através do resultado da função de perda, a rede atualiza os pesos dos filtros de convolução por meio de uma função de optimização para minimizar o erro entre a previsão e a classe original. Esta aprendizagem é feita através do algoritmo retro-propagação, onde se computam as derivadas em cadeia da função de perda e se atualizam os pesos usando essas derivadas parciais, até se minimizar o erro. A atualização dos pesos pode ser realizada em cada amostra que passa pela rede e este algoritmo de optimização designa-se por Gradiente Descendente Estocástico. No entanto, apresenta a desvantagem de tornar o treino ruidoso e pode demorar mais tempo a convergir para o mínimo global. A rede também pode atualizar os pesos após a iteração de todas as amostras de treino, contudo isso implica um grande custo computacional e torna o treino mais moroso. O ideal é atualizar os pesos, a cada iteração de um sub-conjunto de amostras, resultando num custo computacional mais eficiente.

Estes algoritmos de optimização têm como objetivo encontrar os pesos ideais que minimizam o custo, diminuindo o tempo de treino. A eficiência do algoritmo e o controlo da variação dos pesos, depende do valor da taxa de aprendizagem e da variável momento [6] [Qian, 1999] que acelera a aprendizagem, permitindo convergir para o mínimo global que minimiza a perda. Se o valor da taxa de aprendizagem for elevado, o modelo poderá não conseguir atingir o mínimo global, caso seja um valor pequeno, a aprendizagem será mais demorosa [Patterson and Gibson, 2017]. Contudo existem algoritmos de optimização que controlam a taxa de aprendizagem e o momento de forma a tornar a aprendizagem mais rápida,

[6] Do inglês *Momentum*

designados por algoritmos de optimização adaptativos, como o RMSprop [7] e Adam [Kingma and Ba, 2014]. O algoritmo de optimização RMSprop permite reduzir as oscilações e diminuir o tempo de aprendizagem. Esta solução atualiza os pesos com base na média dos valores anteriores do peso ao quadrado. O algoritmo Adam combina as vantagens do RMSprop e do momento, limitando as oscilações na deteção do mínimo local porque adapta a taxa de aprendizagem com base nos pesos, e aproveita as vantagens do momento no processo de aceleração da aprendizagem.

De forma a se verificar e controlar o desempenho da RNC é importante dividir o conjunto de dados em amostras utilizadas para o treino do algoritmo, amostras de validação para verificar se a rede apresenta o desempenho esperado e correto. Por último, após o treino e validação da rede, aplica-se a rede para a classificação de novas amostras nunca observadas para concluir se o modelo consegue classificar as amostras com uma taxa de acerto elevada.

**Regularização:** Ao longo do treino, é importante verificar o comportamento do erro do conjunto de treino e de validação e, consoante o problema, ajustar os híperparâmetros da rede, que estão relacionados com a arquitetura e configuração do treino das RNC, tais como o número de épocas, o número de camadas e de unidades e outras configurações, tais como a seleção da função de ativação, o tamanho do conjunto de amostras por iteração, entre outros. Os híperparâmetros influenciam indiretamente, através do treino, os parâmetros internos do modelo, como por exemplo, os pesos.

O número de parâmetros a serem aprendidos está relacionado com a capacidade de generalização do modelo que influencia o desempenho da rede como se verifica na Figura 2.5. Caso o modelo apresente uma capacidade reduzida, implica uma fraca taxa de acerto na classificação, sendo que o valor de erro de treino será elevado [8], ou seja, o modelo encontra-se em sub-ajuste dos dados [9] e a rede não consegue aprender a classificar corretamente os dados. De forma a evitar este problema,

[7] Referido em www:cs:toronto:edu= tijmen=csc321=slides=lectureslideslec6:pdf
[8] Designado por *bias*
[9] Do inglês *underfit*

a solução consiste em tornar o modelo mais complexo, ou seja, aumentar a sua capacidade. No entanto, esse aumento de capacidade pode implicar numa diferença considerável entre o erro de treino (curva a azul na Figura 2.5) e de validação (curva a verde) [10], o que demonstra que a rede apresenta o problema de sobre-ajuste dos dados [11] representado pela curva verde referente aos dados de validação e em forma de U na Figura 2.5. Este problema ocorre quando a rede aprende a classificar os dados de treino, mas quando se testa a classificação em novas imagens, o valor da taxa de acerto é bastante inferior [Goodfellow et al., 2016].

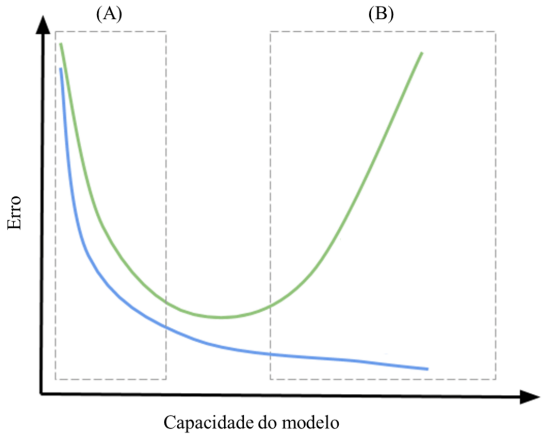



FIGURA 2.5: Desempenho em função da capacidade do modelo. À medida que a capacidade aumenta, o erro de treino e de validação diminui (A), no entanto, a partir de um determinado valor de capacidade, o erro aumenta gerando uma elevada diferença entre o erro de treino e de validação (B). Imagem adaptada de [Mallick, 2019].

De forma a atenuar o sobre-ajuste da rede, são aplicadas técnicas de regularização como a geração de dados sintéticos [12], desativação aleatória de neurónios da rede [13], paragem antecipada do treino, regularização do tipo L1 e L2 e Normalização do valor de entrada [14]. A geração de dados sintéticos permite reduzir a variância do modelo, aplicando-se transformações às imagens durante o treino. Para além

[10] Designado por *variance*
[11] Do inglês *overfit*
[12] Do inglês *data augmentation*
[13] Do inglês *dropout*
[14] Do inglês *batch normalization*

disso, é uma técnica útil para conjuntos de dados de pequenas dimensões permitindo gerar novas imagens a partir das existentes, através de transformações como rotações, ampliação das imagens, entre outras.

A desativação de neurónios [Srivastava et al., 2014] consiste em desligar aleatoriamente um certo número de unidades de uma camada em cada iteração, consoante um valor probabilístico. Assim, a rede é obrigada a aprender novamente, prevenindo a adaptação aos dados de treino. Paragem antecipada consiste em interromper o treino da rede quando o valor da perda nos dados de validação deixa de decrementar, enquanto o de treino continua a diminuir ao longo das épocas. Ou seja, terminar o treino na época em que o modelo entra em sobre-ajuste, quando a diferença associada ao erro entre os dados de treino e de validação ultrapassa um certo valor após a estabilização do valor da perda.

Tanto o método de regularização L1 como o L2, relacionam-se com a alteração da função de perda com o objetivo de minimizá-la [Goodfellow et al., 2016]. A normalização do valor de entrada [15] [Ioffe and Szegedy, 2015] trata-se de uma técnica recente em que se realiza a normalização dos valores resultantes das camadas de ativação anteriores impedindo flutuações extremas dos pesos da rede. Esta técnica permite reduzir o tempo de treino.

**Tipos de treino:** Existem algumas investigações relacionadas com a classificação automática de objetos que recorrem a redes pré-treinadas e que apresentam um valor de taxa de acerto elevado. Essas redes pré-treinadas, por norma, são as redes que se destacaram no concurso de classificação automática de objetos, designado por *ImageNet Large Scale Visual Recognition Challenge* [Patterson and Gibson, 2017]. Existem pelo menos duas formas de utilizar estas redes. A primeira consiste em guardar os valores dos pesos originais das redes e ajustar as últimas camadas convolucionais [16], responsáveis pela deteção do detalhe dos padrões das amostras recebidas. Treina-se somente essas últimas camadas convolucionais para se atualizar os padrões aos novos dados de treino, ou então, alterar somente a configuração da camada *softmax* para o número de classes do problema. Desta

[15] Do inglês *batch normalization*
[16] Do inglês *fine-tune*

forma, reduz-se o tempo de treino. A segunda técnica consiste em re-treinar toda a rede pré-treinada, no entanto, implica a utilização de conjuntos de dados de treino robustos para que não ocorra sobre-ajuste aos dados. O uso de redes pré-treinadas permite convergir de forma mais rápida, no entanto, não é útil utilizar os pesos das redes pré-treinadas em conjuntos de dados que não sejam semelhantes do conjunto usado para o treino das mesmas. Também não é aconselhável treinar as redes pré-treinadas com conjuntos de amostras reduzidos, porque poderá ocorrer o sobre-ajuste aos dados devido à elevada capacidade dos modelos. Outra forma de aprendizagem, consiste na construção de uma RNC de origem e no seu treino, modificando a rede de forma a que esta atinja o desempenho objetivo.

**Ferramentas:** O modelo proposto nesta dissertação foi construido utilizando os softwares keras [Chollet, 2019] e tensorflow [Google, 2019], que são bibliotecas de código aberto para a construção e treino de redes neuronais e outros algoritmos de aprendizagem automática. Para o processamento das imagens foi utilizada a biblioteca OpenCV [Intel, 2019], que é útil para aplicações da área da visão computacional. Os modelos foram treinados utilizando a GPU Tesla k80 da ferramenta Colaboratory da Google [Google, 2018].


FIGURA 2.6: Diagrama com a relação entre as ferramentas usadas. O OpenCV é utilizado para ler imagens e modificá-las para serem recebidas pelos modelos desenvolvidos através da biblioteca Keras (1) que recorre ao Tensorflow para realizar os cálculos de baixo nível (2).

# Capítulo 3

# Estado da Arte

Neste capítulo é apresentado trabalho anterior que estuda a deteção e reconhecimento automático de objetos em imagens aéreas, através de técnicas baseadas em aprendizagem automática. O primeiro tópico relaciona-se com a classificação de imagens por meio de aprendizagem automática tradicional, ou seja, modelos não baseados em redes profundas. No segundo, são referidos estudos sobre a deteção de objetos e a segmentação de imagens recorrendo a técnicas de aprendizagem profunda, que se trata de uma sub-área da aprendizagem automática. Na última secção estão presentes artigos relacionados com a classificação e monitorização de espécies da flora utilizando RNC.

## 3.1 Aprendizagem Automática na classificação de imagens aéreas

A deteção remota em conjunto com algoritmos de aprendizagem automática clássicos, permite a classificação automática de imagens de grande escala. Esta evolução tornou-se útil para o estudo da ocupação do território [Mendes and Dal Poz, 2011], distinção de zonas [Tudorache et al., 2017], monitorização e construção de mapas de vegetação. Estes mapas tornam-se úteis para o conhecimento da distribuição de espécies numa dada área e para o controlo de espécies invasoras [Martins,

2012, Gil et al., 2013, Pande-Chhetri et al., 2017, Lopatin et al., 2019, Paz-Kagan et al., 2019].

Em Mendes e Dal Poz [Mendes and Dal Poz, 2011] recorreu-se a redes neuronais artificiais para a classificação da ocupação do território em imagens aéreas, distribuída por seis classes. Para tal utilizou-se dois conjuntos de dados diferentes: um que contém imagens RGB com dados radiométricos e o segundo, para além de usar os dados radiométricos, apresenta um Modelo Digital de Superfície Normalizado com uma representação da elevação dos objetos numa superfície plana. Estes dois tipos de dados foram aplicados à rede neuronal respetiva e calcularam o valor de Kappa ($\kappa$), que fornece informação sobre o quão próximas as classes previstas estão das verdadeiras, para verificar o nível de concordância entre a classificação do modelo e a classificação de referência. Concluíram que o conjunto de amostras com informação extra de dados radiométricos permite um aumento do desempenho do modelo, atingindo-se um valor de $\kappa$ de 0.86.

Tudorache et al. [Tudorache et al., 2017] desenvolveram um método recorrendo a uma RNA para classificar as áreas como sendo vegetação ou zonas de inundação através das características da textura e do Histograma de Gradientes Orientados (HOG). A metodologia proposta envolve o treino de quatro RNA, sendo que duas são usadas para a identificação de vegetação e outras duas aplicadas para a deteção de zonas de inundação. Em cada par de redes neuronais, uma é alimentada com as características de textura e a outra com o Histograma de Gradientes Orientados, classificando cada fragmento da imagem como vegetação ou não, ou como zona de inundação. Os autores concluíram que para diminuir a taxa de classes não identificadas corretamente, seria melhor usar um conjunto de dados mais elaborado, o que pode implicar aumentar o tempo computacional e a taxa de falsos positivos devido à existência de mais elementos com texturas semelhantes.

Em relação à monitorização e identificação de espécies invasoras, Martins [Martins, 2012] estudou a existência de áreas invadidas pela espécie *Acacia dealbata* através de imagens de satélite multi-espectral ASTER. O autor comparou três técnicas de

aprendizagem automática supervisionada para a classificação das imagens, nomeadamente um classificador de Máxima Verosimilhança, um classificador baseado em SVM e um classificador baseado numa RNA. Verificou-se que o classificador de Máxima Verosimilhança apresentou melhores resultados na classificação e na monitorização desta espécie, com valor de precisão global de 73% e um valor de $\kappa$ de 0.7.

De forma semelhante à investigação anterior, Lopatin et al. [Lopatin et al., 2019] realizam o mapeamento das espécies invasoras *Acacia dealbata*, *Ulex europaeus* e *Pinus radiata* em imagens aéreas capturadas por VANTs. Para tal recorreram ao classificador Máxima Entropia variando as características utilizadas para o seu treino e teste, avaliando-o com base na área sob a curva, a métrica $\kappa$, sensibilidade e especificidade do modelo. Concluíram que o uso da informação da sombra afeta o desempenho do modelo.

Para a monitorização da espécie invasora *Acacia Longifolia*, Sá et al. [de Sá et al., 2018] utilizaram o classificador florestas aleatórias [1] para a criação de um mapa com a distribuição da floração desta espécie, atingindo um desempenho na identificação superior a 96%. Com este estudo concluíram que se torna possível controlar e monitorizar a evolução da floração da espécie, após a aplicação do agente de controlo biológico.

A utilização das técnicas de aprendizagem automática referidas implica a preparação prévia dos dados, selecionando corretamente as características importantes para treinar os classificadores. De forma a melhorar o desempenho do modelo, é necessário aplicar informações adicionais, como foi demonstrado na investigação anterior [Tudorache et al., 2017]. No entanto, com a aplicação de aprendizagem profunda este pré-processamento de informação é descartado e portanto, neste projeto de dissertação recorreu-se a modelo de aprendizagem profunda para a classifcação automática de imagens. Nas próximas secções serão referidos trabalhos relacionados com a classificação de imagens aéreas recorrendo a redes neuronais convolucionais.

[1] Do inglês *Random Forest*

## 3.2 Aprendizagem Profunda na classificação de imagens aéreas

Nesta dissertação recorreu-se a redes neuronais convolucionais para a aprendizagem automática de padrões em imagens, distinguindo-as das várias classes. A classificação e a deteção de objetos em imagens, são duas das aplicações das redes neuronais convolucionais. A classificação consiste em identificar a classe a que pertence um objeto numa dada imagem resultando num vetor com a probabilidade de pertencer a cada classe, enquanto que na deteção, o objetivo é percorrer a imagem com uma janela deslizante, que será recebida pela rede para realizar a previsão da classe e retorna a localização do objeto através de informações da caixa delimitadora gerada. De seguida, encontram-se descritos estudos relacionados com a classificação e deteção de objetos em imagens aéreas.

### 3.2.1 Classificação e Deteção de objetos

As redes neuronais convolucionais, tal como os algortimos de aprendizagem automática, são utéis para o estudo da ocupação do território. Em Castelluccio et al. [Castelluccio et al., 2015] estudaram-se dois tipos de redes neuronais convolucionais que se destacaram no concurso de classificação de imagens *ImageNet Large Scale Visual Recognition Challenge*, as redes CaffeNet e GoogLeNet, para classificar imagens aéreas provenientes de satélites na banda dos infra-vermelhos e imagens óticas. A rede GoogLeNet apresenta uma técnica que permite reduzir o custo computacional através de camadas compostas por conjuntos de filtros de convolução de dimensões diferentes [2] representados na Figura 3.1, ao invés de filtros com a mesma dimensão por camada. Estas redes foram treinadas aplicando três métodos diferentes: no primeiro, treinaram de raiz usando os conjuntos de dados, no segundo, testaram numa rede pré-treinada e adaptaram algumas camadas ao problema (*fine-tuning*) e no terceiro, usaram os resultados da penúltima camada pré-treinada como vetor de características, que será lido pela camada

[2]Designados por módulo *inception*

*softmax* que realiza a classificação, sendo que só a última camada completamente conectada é que é treinada. Para o conjunto de imagens aéreas óticas, a segunda técnica apresentou melhores resultados nas duas redes neuronais convolucionais, atingindo uma precisão de 95.48% e de 97.10%, para CaffeNet e GoogLeNet, respetivamente. Para o conjuntos de imagens na banda dos infra-vermelhos, o melhor resultado de desempenho obtido foi de 91.8%, na experiência onde se aplicou a primeira técnica utilizando a rede GoogLeNet. Em geral a rede GoogLeNet apresentou melhores resultados devido ao elevado número de amostras por cada classe. No entanto em imagens de classes próximas, as classificações apresentaram piores resultados, mas os autores afirmam que uma fusão adequada dos resultados de CaffeNet e GoogLeNet ajudaria a remover os erros.

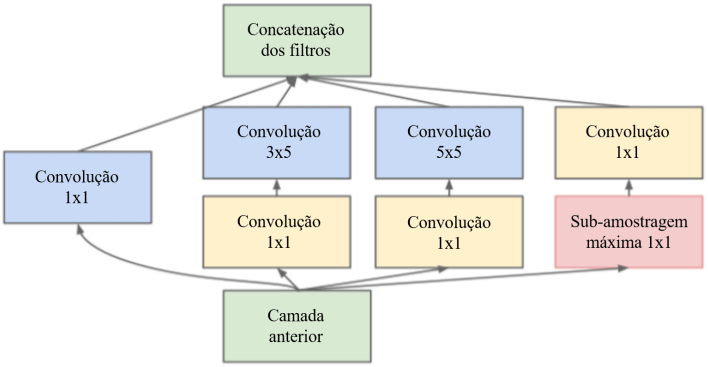



FIGURA 3.1: Composição de uma camada presente na rede GoogLeNet. Neste tipo de camada são realizadas convoluções com três filtros de dimensões diferentes no mapa de características da camada anterior. Posteriormente, os resultados são concatenados para serem recebidos pela camada seguinte.

Em Sameen et al. [Sameen et al., 2018] foi desenvolvida uma rede neuronal convolucional para a classificação da ocupação do território em sete classes. Os autores realizaram uma análise de sensibilidade para estudar os híperparâmetros mais adequados à rede. Aplicaram técnicas de regularização, como a desativação aleatória de unidades e normalização dos dados, testando individualmente e em conjunto o desempenho da rede. Compararam com os resultados obtidos sem a utilização destas técnicas e concluíram que a união dos dois tipos de regularização permite melhorar o desempenho da rede, reduzindo o sobre-ajuste aos dados. Os autores alertam para as vantagens e desvantagens do aumento da dimensão e do número de filtros de convolução. Concluíram que o aumento da dimensão do filtro permite

melhorar o desempenho da rede mas, no entanto, implica num aumento do custo computacional. Da mesma forma, apesar do aumento do número de filtros ser benéfico para o modelo, este poderá atingir o sobre-ajuste aos dados de treino devido à complexidade atingida. Em relação à função de ativação, a função ReLU foi a que se destacou, permitindo diminuir o tempo de treino. O algoritmo de Optimização NADAM [Dozat, 2016] permitiu atingir o melhor valor de desempenho e provém da optimização do algoritmo Adam. Após a análise de sensibilidade e de regularização, a rede final é composta por uma camada de convolução acompanhada de sub-amostragem máxima, normalização dos dados com uma probabilidade de 50% de desativação de neurónios e duas camadas completamente conectadas com a função de ativação *softmax*. Conseguiram alcançar 97.3% de precisão geral, 96.5% de média e 0.967 da métrica $\kappa$.

Sevo e Avramović [Ševo and Avramović, 2016] ajustaram a rede neuronal convolucional GoogLeNet com o fim de detetar e classificar objetos e regiões em cinco classes, com base em imagens aéreas. Os autores realizaram duas experiências de forma a perceber o efeito do uso de diferentes algoritmos do gradiente descendente. Testaram com o gradiente descendente estocástico e com o algoritmo gradiente adaptativo, em que o valor da taxa de aprendizagem se adapta consoante os parâmetros. Concluíram, em comparação com outros estudos existentes, que o segundo método apresenta o melhor valor de precisão na classificação, cerca de 98%.

As RNC também são utilizadas para a deteção de determinados objetos. Em Ammour et al. [Ammour et al., 2017] é apresentada uma solução automática para detetar carros em imagens obtidas por veículos aéreos não tripulados combinando métodos de segmentação, como o algoritmo *Mean-shift* para extrair regiões com probabilidade de conterem um carro e uma rede neuronal convolucional para extrair as características das regiões obtidas anteriormente, de forma a serem classificadas como carro pelo classificador SVM. Aplicaram aumento sintético dos dados para aumentar o número de imagens de treino. Os autores conseguiram diminuir

o tempo computacional devido à aplicação do algoritmo de segmentação, que permitiu reduzir o espaço de pesquisa e melhorar a deteção de objetos através de operações morfológicas.

Continuando na área da deteção de carros, Tang et al. [Tang et al., 2017] propuseram um modelo adaptado da rede neuronal convolucional Faster R-CNN [Ren et al., 2015], para detetar veículos em imagens aéreas e estimar a sua orientação. O modelo divide-se em dois submodelos com funções diferentes. O primeiro designa-se por VPN, *Vehicle Proposal Networks* e consiste numa arquitetura completamente convolucional (RCC), ou seja, uma rede sem camadas completamente conectadas, cujo objetivo é gerar regiões candidatas a conterem veículos com diferentes resoluções e escalas. Posteriormente, as imagens das regiões candidatas são recebidas por outra rede composta por camadas convolucionais e por uma camada de sub-amostragem de região de interesse de forma a tornar cada uma com dimensão fixa para serem recebidas pelas camadas completamente conectadas. Por fim, usa-se a função *softmax* para classificar as orientações dos veículos. Apesar do modelo ser rápido e apresentar bons resultados na deteção em relação a outros algoritmos, ele exibe um pior desempenho na deteção de veículos nas imagens de satélite em larga escala. Para além disso, existem veículos que não são detetados e falsos positivos, ou seja, objetos considerados como veículos, mas que não pertencem a essa classe. Este modelo expressa mais uma técnica de deteção de objetos e segmentação de áreas de interesse.

Na investigação de Radovic et al. [Radovic et al., 2017] realizou-se a deteção de aviões em imagens aéreas através de uma RNC baseada no algoritmo de deteção YOLO [Redmon et al., 2016]. Este algoritmo devolve um vetor com informações sobre a classe prevista, a probabilidade de pertencer à classe em questão e as coordenadas da caixa delimitadora no espaço da imagem como se verifica pela Figura 3.2. Durante o treino, testaram diferentes valores dos hiperparâmetros e recorreram à geração de dados sintéticos e desativação aleatória de neurónios. Após o treino, testaram a rede para a deteção de aviões atingindo uma precisão de 97.5%. Por fim, aplicaram o algoritmo para realizar a deteção em tempo real através de um VANT, alcançando-se uma taxa de acerto de 86%. Os autores concluiram que

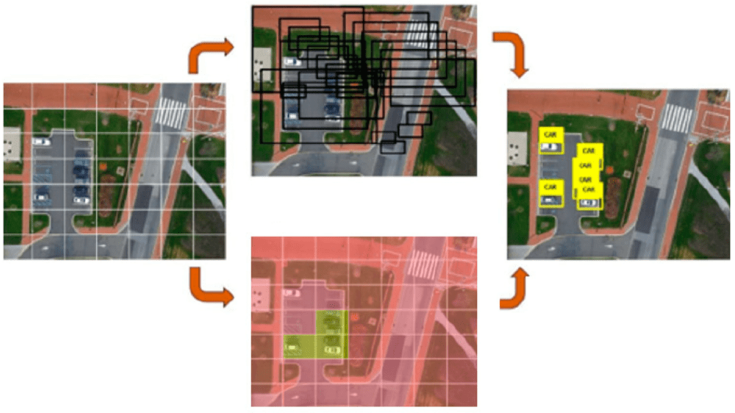


FIGURA 3.2: Deteção de objetos através do algoritmo YOLO. Neste algoritmo, divide-se a imagem em sub-conjuntos de píxeis com igual dimensão em que para cada sub-conjunto se prevê as caixas delimitadoras necessárias para abranger todos os objetos e o mapa de probabilidades das classes. Figura retirada do artigo [Radovic et al., 2017].

a proposta desenvolvida é uma abordagem candidata para a deteção de objetos em tempo real usando veículos aéreos não tripulados, já que permite detetar os objetos oclusos por outros, para além da sua alta taxa de acerto.

Deng et al. [Deng et al., 2018] desenvolveram um método para a deteção de objetos de várias escalas usando duas redes, a rede de propostas de objetos de várias escalas designada por MS-OPN e a rede de deteção precisa de objetos AODN. No entanto, existe uma RNC composta por um conjunto de filtros de convolução de dimensões diferentes [3], substituindo filtros com a mesma dimensão por camada. Os resultados de cada filtro de convolução são concatenados e recebidos pelas duas redes anteriores. De forma a que a rede MS-OPN consiga gerar regiões de objetos com filtros de dimensão diferentes, são usadas três janelas deslizantes de dimensões diferentes em cada um dos mapas de características. Posteriormente, a rede AODN recebe a imagem com a previsão das caixas delimitadoras geradas anteriormente. Através da combinação das informações obtidas pelas camadas de várias resoluções, é construído um mapa de características que permite obter uma melhor precisão na deteção de objetos. Testaram o algoritmo em quatro conjuntos de dados de imagens aéreas para a deteção de diferentes classes, aplicando aumento

[3]Designados por módulo *inception*

sintético dos dados durante o treino. Compararam os resultados com outros métodos, concluindo que o método proposto permite detetar os objetos a várias escalas e apresentou melhor desempenho na identificação de objetos de pequeno tamanho em comparação com os outros métodos.

Por fim, Liu et al. [Liu et al., 2018] avaliaram e compararam diferentes algoritmos de aprendizagem profunda para a deteção e segmentação de imagens aéreas no reconhecimento de pássaros. Em relação à deteção de pássaros analisaram o método Single Shot MultiBox Detector (SSD) [Liu et al., 2016] que localiza e classifica o objeto numa única passagem na rede neuronal convolucional adaptada da arquitetura VGG-16 [Simonyan and Zisserman, 2014]. O desempenho deste algoritmo na deteção de objetos mais pequenos decresce e por isso autores optaram por descartá-lo. O Single Stage Headless (SSH) [Najibi et al., 2017] é outro algoritmo de deteção construído com base na arquitetura VGG-16 e comporta-se melhor na deteção de objetos pequenos, pois adiciona mais informação de contexto que extrai das camadas convolucionais, tal como o Tiny Face [Hu and Ramanan, 2017]. Por último, analisaram o algoritmo You Only Look Once (YOLO) versão 2 [Redmon et al., 2016], em que se usa normalização dos dados para estabilizar a rede e melhorar o desempenho. Em relação aos métodos de segmentação, observaram o comportamento da arquitetura U-Net [Ronneberger et al., 2015] e Mask R-CNN [He et al., 2017]. Testaram os vários algoritmos em diferentes conjuntos de dados e concluíram que para detetar pássaros em imagens com um fundo que dificulta o reconhecimento dos mesmos, por exemplo, com vegetação abundante, o algoritmo Tiny Face apresenta melhor desempenho. Para imagens mais simples, o SSH foi o algoritmo que mostrou ser mais eficiente na deteção. Em relação à segmentação, a rede U-Net realçou melhor desempenho em comparação com o método Mask R-CNN. Este trabalho anterior utilizou imagens com abundância de informação, tal como as que serão usadas na solução desenvolvida nesta dissertação e através dos resultados é possível perceber que algoritmo usar para detetar objetos de pequenas dimensões consoante o nível de informação presente nas imagens aéreas.

### 3.2.2 Segmentação de imagens

As redes neuronais profundas permitem a segmentação de imagens, isto é, classificação ao nível do pixel que resulta num mapa com áreas de objetos de classes diferentes, detetados na imagem original. Neste tipo de classificação, usualmente, converte-se uma rede pré-treinada para uma arquitetura de rede completamente convolucional (RCC) em que se substituem as camadas completamente conectadas por convolucionais como se ilustra na Figura 3.3. Estas redes recebem imagens de alta resolução e devolvem um mapa de classificação com a mesma dimensão que a imagem original.

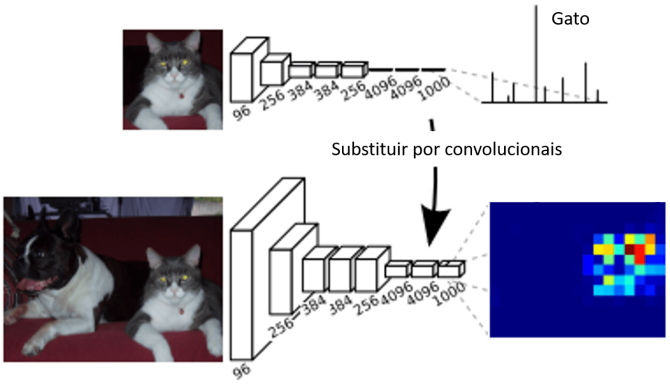



FIGURA 3.3: Substituição das camadas conectadas por camadas convolucionais para a segmentação. Imagem retirada do artigo dos autores [Long et al., 2015].

Existem várias formas para se obter um mapa de classificação com a mesma dimensão que a imagem recebida pela rede. O autor Long et al. [Long et al., 2015] optou por aplicar uma operação de deconvolução, ou seja, operação inversa à operação de convolução para aumentar a resolução [4] do mapa de características com filtros de deconvolução que são ajustados ao longo da aprendizagem da rede. Adicionaram módulos residuais [5], que representam ligações diretas entre camadas não consecutivas, para lidar com o problema de perda de resolução devido às camadas de sub-amostragem. Esta técnica combina mapas de características de camadas diferentes. No estudo desenvolvido por Huang et al. [Huang et al., 2018]

[4] Designado por *upsampling*
[5] Designado por módulos *skip*

verificou-se a utilização deste tipo de arquitetura para a segmentação de imagens aéreas.

O modelo SegNet [Badrinarayanan et al., 2017] representa uma alternativa de se obter um mapa de classificação com dimensão da imagem original. Este modelo é composto por uma arquitetura codificador-descodificador ilustrada na Figura 3.4. A fase do codificador é constituída por uma rede completamente convolucional para a deteção de características. De forma a se obter um mapa de classificação com a dimensão da imagem original, o descodificador utiliza as informações obtidas pelos índices das camadas de sub-amostragem do codificador não sendo necessário aprender os filtros de deconvolução como nas RCC. Cada camada do descodificador está associado a uma camada no codificador e através dos índices de sub-amostragem guardados, gera-se um mapa de características no descodificador com as mesmas dimensões que o mapa respetivo na camada do codificador. Os mapas de características resultantes são preenchidos pelos valores de sub-amostragem máxima nas posições correspondentes aos índices guardados e as restantes posições são preenchidas por zeros. Este mapa de características resultante sofre operações de convolução com os filtros do descodificador que são aprendidos durante o treino para a geração de mapas de características mais densos. O resultado da camada do descodificador é recebido por uma camada *softmax* para a classificação ao nível do pixel. Este modelo apresenta desativação aleatória de neurónios e aumento sintético dos dados, como métodos de regularização.

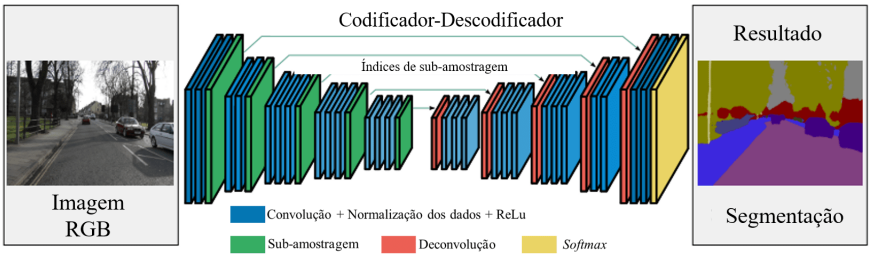



FIGURA 3.4: Arquitetura da rede de segmentação de imagens SegNet. Codificador é responsável pela realização da convolução e o descoficador realiza a desconvolução para se obter um mapa de classificação com dimensão igual à imagem original. A Figura foi adaptada do artigo [Badrinarayanan et al., 2017].

O modelo proposto por Audebert et al. [Audebert et al., 2017] designa-se por *Segment-before-Detect*, que combina a segmentação e a deteção, recorrendo à rede SegNet para segmentar a imagem. Depois da segmentação, realiza-se a extração dos componentes conectados criando a máscara de veículos e aplica-se operações morfológicas para remover o ruído e falsos positivos, aumentando a precisão. Através das caixas delimitadoras obtém-se os veículos que são recebidos por uma rede neuronal convolucional pré-treinada e ajustada para classificar os tipos de veículos detetados. Os autores compararam os resultados com outros métodos e concluíram que aplicar segmentação antes da deteção de veículos melhora o desempenho do modelo, alcançando-se valores máximos de precisão e sensibilidade de 90% e 84%, respetivamente. Esta investigação usou desativação aleatória dos neurónios, aumento sintético dos dados e renormalização dos dados como técnicas de regularização.

Existe outro estudo [Ham et al., 2018] que realiza a segmentação das imagens aéreas, obtidas por VANTs, para a deteção de construções não registadas usando redes neuronais deconvolucionais [Noh et al., 2015] que se trata de outra alternativa para realizar a segmentação. Nestas redes neuronais deconvolucionais, as camadas completamente conectadas permanecem na rede, como se verifica na arquitetura da rede representada na Figura 3.5, ao contrário das RCC e da arquitetura SegNet. O processo para obter um resultado com a mesma dimensão que a imagem original é o mesmo que na rede SegNet. A rede desenvolvida pelos autores segmenta as imagens classificando cada pixel como edifício ou não, obtendo-se um mapa de probabilidades. Depois subtraem o mapa de previsões com informações do mapa digital e os píxeis que restam da operação são considerados como edificações não registadas. De forma a remover o ruído causado pelas previsões, por não coincidirem completamente no formato das construções, os autores aplicam operações morfológicas. Após a avaliação do modelo, concluíram que este apresenta pequenas limitações na segmentação de edifícios altos e na deteção de falsos positivos. No entanto, é um modelo que pode ser utilizado para a deteção e monitorização de construções ilegais.

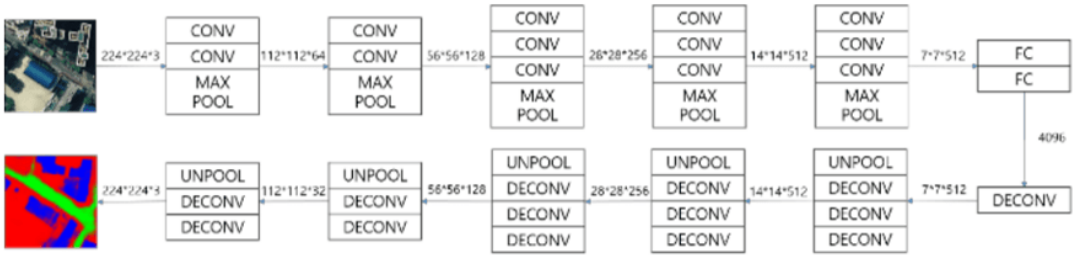



FIGURA 3.5: Arquitetura da rede deconvolucional utilizada no artigo em questão. Como se pode verificar, nesta redes não se substituem as camadas conectadas por convolucionais.

Maggiori et al. [Maggiori et al., 2017] começaram por analisar as arquiteturas que realizam a segmentação semântica de imagens, como a rede neuronal convolucional com camadas de dilatação, módulos residuais e deconvolucional. Através das conclusões da análise anterior, desenvolveram uma nova abordagem, baseada nas vantagens da rede com módulos residuais. Este modelo extraí várias características provenientes das camadas intermédias de uma rede completamente convolucional que serão combinadas através de uma rede neuronal simples, que aprende como combinar essas características, originado um mapa final de classificação píxel a píxel. Como técnicas de regularização usaram aumento sintético dos dados, regularização L2 e normalização dos dados. Para treinar, aplicaram dois conjuntos de dados diferentes e concluíram que a utilização de características de baixa resolução, incrementa a precisão de classificação em comparação com outras redes, verificando uma precisão geral de 88%.

Sun et al. [Sun et al., 2018], realizaram a segmentação semântica de imagens aéreas, combinando uma RNC multi-filtro, que recebe a concatenação de imagens de alta resolução com informação proveniente da técnica de deteção remota Lidar, com uma segmentação multi-resolução, para delimitar os objetos. Usaram a rede SegNet com normalização dos dados para extrair as características da imagem recebida pela rede. Os autores recorrem a três redes SegNet em paralelo, sendo que os mapas de características extraídos por um filtro com dimensão diferente para cada rede, são recebidos individualmente por um classificador *softmax* para se realizar o cálculo da função de perda durante o treino. Treinaram e testaram o modelo com dois conjuntos de dados que diferem na existência ou não da banda

de radiação infravermelha na sua composição. O primeiro conjunto de dados é composto por imagens com quatro canais, vermelho, verde, azul e com informação de radiação infravermelha e o segundo, contém imagens com informação RGB. Os mapas resultantes das três SegNet são, posteriormente, concatenados de forma a serem recebidos por outro classificador *softmax*, que compara as classes previstas com as verdadeiras. O algoritmo de segmentação de multi-resolução é utilizado para reduzir o ruído obtido na classificação das imagens. Após as experiências, concluíram que esta metodologia permite tornar os limites da segmentação mais suaves, permitindo obter melhores resultados de precisão, cerca de 90% para o primeiro conjunto de dados e de 88% para o segundo, em comparação com outros métodos.

Nesta secção foram introduzidas os tipos de classificação que as redes neuronais convolucionais permitem realizar e foram abordadas investigações que recorreram a métodos de aprendizagem profunda para realizar a deteção, reconhecimento e segmentação automática de objetos em imagens aéreas. Foram abordados temas relacionados com a deteção de carros, a ocupação do território, entre outros. O tema desta dissertação está relacionado com a deteção de espécies invasoras e por isso, na próxima secção serão referidos artigos que abordam a deteção de espécies da flora.

## 3.3 Monitorização e deteção de vegetação em imagens aéreas

Nesta secção, serão referidas investigações relacionadas com o tema do projeto, nomeadamente a monitorização e deteção de vegetação em imagens aéreas capturadas por VANTS que recorrem a técnicas de aprendizagem profunda. Em Wuang et al. [Wang et al., 2019], os autores desenvolveram um algoritmo de segmentação de super-píxel que consiste em segmentar a imagem em conjunto de píxeis que representam o mesmo objeto. Recorreram a uma rede neuronal convolucional para segmentar uma imagem aérea consoante a existência da espécie *Ulva prolifera*.

Utilizam, então, o algoritmo de segmentação de super-pixel designado por SEED [Van den Bergh et al., 2012] para dividir a imagem em conjuntos de píxeis que pertencem ao mesmo objeto, através da cor. Posteriormente, cortam a imagem em diferentes partes em função do centro de cada super-pixel e escalam-nos para serem recebidos pela RNC, ALexNet, para a classificação binária como contendo a espécie em estudo ou não. Após cada conjunto de píxeis ter sido classificado atualizam a matriz gerada com o valor da classe de cada super-pixel da imagem, ou seja, agregam todos os conjuntos classificados, originando a imagem inicial segmentada. O método desenvolvido apresentou um desempenho superior à RCC de comparação e concluíram que pode ser usado para distinguir os diferentes tipos da alga *Ulva prolifera*.

Em [Fan et al., 2018] é proposto um algoritmo para a deteção de plantas de tabaco em imagens aéreas obtidas por VANTs. A solução apresentada começa por extrair as regiões candidatas a conterem a planta do tabaco através de operações morfológicas e utilizam um algoritmo de segmentação para segmentar as regiões com a planta de interesse, através da intensidade das cores. Cada segmento obtido anteriormente é classificado, como região que contém planta do tabaco ou não, através de uma rede neuronal convolucional. Os autores realizaram três experiências, sendo que em todas foram testadas duas RNC que diferem na existência ou não de camadas de sub-amostragem na sua arquitetura. Na primeira experiência usam as RNC exclusivamente para a classificação. Na segunda, utilizam classificadores clássicos como o vizinho mais próximos [6], SVM e o algoritmo de florestas aleatórias que recebem as características obtidas pelas RNC. Por fim, numa terceira experiência avaliaram a influência da variação do número de exemplos de treino no desempenho da rede. Concluíram que a RNC com camadas de sub-amostragem permite alcançar melhor desempenho na identificação de plantas de tabaco do que a segunda rede. Para além disso, os resultados da experiência em que usam os classificadores de aprendizagem automática mostram que as RNC conseguem extrair as características importantes para a classificação de imagens.

[6] Do inglês *Nearest neighbor*

Os autores comprovaram, mais uma vez, que o número de amostras do conjunto de dados de treino influencia o desempenho da rede neuronal convolucional.

Onishi and Ise [Onishi and Ise, 2018], desenvolveram um algoritmo com base na rede GoogLeNet para classificar automaticamente imagens aéreas consoante sete espécies de árvores. Realizaram duas experiências aplicando ou não geração sintética de dados nos conjuntos de imagens de treino e validação. Comparam os resultados dos dois ensaios e verificaram que os valores de precisão de algumas classes rondavam os 0%, já que o número de amostras de três classes era reduzido em comparação com o número de imagens das outras espécies. Na segunda experiência recorreram à geração sintética de dados e conseguiram reduzir o problema da existência de falsos negativos das três classes problemáticas, atingindo um valor de 89% de classes classificadas corretamente. Pode-se concluir que a técnica de aumento dos dados permite melhorar o desempenho de uma rede, principalmente se existir um elevado desequilíbrio de número de amostras por classes.

Em Sa et al. [Sa et al., 2017] também realizaram a segmentação de imagens aéreas multi-espectrais com três canais, em três classes: áreas de cultivo, zonas com ervas que não pertencem à plantação e outros, recorrendo à rede SegNet. Realizaram seis experiências onde variaram o número de canais da informação recebida pela rede e o modo de treino, treinando de raiz ou utilizando a rede pré-treinada. Verificaram que esta última abordagem não apresentou resultados com impacto significativo comparando com os obtidos para a deteção de outros objetos, devido ao reduzido número de amostras relativamente ao usado para treinar a rede pré-treinada. Concluíram, com base nas imagens segmentadas das diferentes redes, que os píxeis que pertencem à classe área de cultivo são por vezes classificados erradamente como ervas que não pertencem à plantação, devido à existência de vários tipos de ervas e à dificuldade em distinguir estas duas classes na fase inicial do desenvolvimento. Os mesmo autores realizaram um novo estudo em [Sa et al., 2018], em que usam mapas multi-espectrais com nove canais de grande escala e uma janela deslizante com as dimensões da imagem recebida pela rede base desenvolvida no estudo anterior. Esta janela deslizante percorre todo o mapa, sendo que o seu conteúdo é recebido pela rede que realiza a segmentação do conjunto de píxeis que

recebe, juntando-se os resultados das várias classificações. Testaram diferentes configurações da rede com imagens de nove canais, alcançando um valor máximo de precisão na segmentação das três classes, superior ao valor de precisão obtido no estudo anterior.

Huang et al. [Huang et al., 2018] recorreram a redes completamente convolucionais para a segmentação de imagens aéreas obtidas por VANTs, separando os píxeis em três classes: planta do arroz, ervas selvagens e outros. Compararam as redes AlexNet, VGGNet e GoogLeNet, substituindo as camadas completamente conectadas por convolucionais, de forma a que estas representassem uma rede completamente convolucional. Posteriormente, modificaram a rede VGGNet, que apresentou melhor desempenho, aplicando-lhe uma camada deconvolucional para converter o resultado para a dimensão da imagem original. Adicionalmente propõem uma arquitetura com módulos de ligação direta para camadas de convolução não consecutivas, de forma a compensar a perda de resolução resultante das camadas de sub-amostragem, combinando o mapa de característica de duas camadas com resoluções diferentes. Concluíram que a utilização de redes completamente convolucionais combinadas com módulos residuais, permitem incrementar os valores de precisão da deteção e classificação por classe e ultrapassar o desempenho de outros métodos de segmentação.

Os autores Safonova et al. [Safonova et al., 2019] recorreram a uma rede neuronal convolucional para a identificação dos estágios de saúde de árvores invadidas por uma espécie invasora da fauna. A rede é composta por seis camadas, com a aplicação de desativação aleatória de neurónios e geração sintética de dados para o controlo do sobre-ajuste aos dados. A rede conseguiu distinguir os quatro estágios de saúde com as seguintes taxas de precisão: 92.75%, 89.86%, 89.66% e 88.89%. Os autores compararam a abordagem desenvolvida com outros classificadores aplicados em imagens aéreas e verificaram que a sua rede apresenta um desempenho superior na deteção do estágio de saúde das árvores.

Através da análise de estudos enquadrados na área da agricultura, os autores de

[Kamilaris and Prenafeta-Boldú, 2018] concluíram que os algoritmos de aprendizagem automática profunda são, maioritariamente, utilizados para a classificação da ocupação do território [Castelluccio et al., 2015, Ham et al., 2018, Ševo and Avramović, 2016, Deng et al., 2018, Maggiori et al., 2017, Sun et al., 2018], monitorização e classificação de plantações agrícolas [Castelluccio et al., 2015, Huang et al., 2018, Sa et al., 2017, Sa et al., 2018, Wang et al., 2019] e contagem de frutas/árvores ou animais [Liu et al., 2018, Gray et al., 2019]. A deteção e segmentação de automóveis trata-se, também, de um assunto estudado [Audebert et al., 2017, Tang et al., 2017, Ammour et al., 2017]. No entanto, a deteção e classificação de espécies invasoras da flora em imagens aéreas capturadas por VANTs, é uma área inexistente com redes neuronais convolucionais.

Nesta dissertação o objetivo é o reconhecimento da espécie *Acacia Longifolia*. Este assunto foi estudado pelos autores de [de Sá et al., 2018], que utilizaram o algoritmo florestas aleatórias para realizar o mapeamento desta espécie. No entanto, como se trata de um classificador clássico é necessária a extração prévia de características para o treino do classificador. Neste caso foi utilizada a cor, mas poderão existir outras características que poderão tornar mais eficiente o reconhecimento da espécie. Com as redes neuronais convolucionais, as características são descobertas pelo modelo ao longo do treino e são as que representam melhor as várias classes a serem aprendidas, sendo mais eficazes que a seleção manual das características. Deste modo, torna-se importante validar a eficácia do uso de redes neuronais convolucionais para detetar a existência de espécies invasoras em imagens aéreas para o posterior desenvolvimento de ferramentas para a deteção e reconhecimento destas espécies. Este documento preenche esta lacuna aplicando com sucesso RNCs no reconhecimento da espécie *Acacia Longifolia*.

# Capítulo 4

# Arquitetura desenvolvida

Neste trabalho, desenvolveu-se uma arquitetura para o reconhecimento de espécies invasoras, nomeadamente a espécie *Acacia Longifolia*. A abordagem seguida, representada resumidamente na Figura 4.1, é composta por uma primeira etapa relacionada com a aquisição e processamento de imagens aéreas através de um VANT. As imagens de grande escala são depois divididas segundo sub-imagens de 200 x 200 píxeis e, posteriormente, cada sub-imagem é agrupada na sua classe correspondente [1]. No final, dividem-se as amostras em conjunto de dados de treino, validação e teste. De seguida é aplicada a classificação e reconhecimento automático da espécie invasora *Acacia Longifolia* nas imagens aéreas anteriores. Nesta etapa são aplicadas duas redes neuronais, previamente treinadas, que diferem no tipo de classificação, sendo uma treinada para a classificação binária, Acacia L. e Não Acacia L. e outra para a multi-classificação segundo nove classes: a classe Acacia L., Outras amarelas, Sobreiro, Vegetação, Estrada, Restos de árvores, Outros componentes, Pequenas árvores e Pinheiros. Nesta última rede obtém-se o seu desempenho para a classificação binária. Após o treino, realiza-se a segmentação de imagens de alta resolução capturadas pelo VANT através da informação das classes previstas pelas RNC desenvolvidas. A segmentação é realizada recorrendo ou a uma grelha regular cuja janela deslizante se desloca a cada sub-conjunto de 200 x 200 pixeis ou a uma janela deslizante com deslocamento de um pixel de cada

[1] Esta etapa foi fornecida pela empresa IntRoSys S.A.

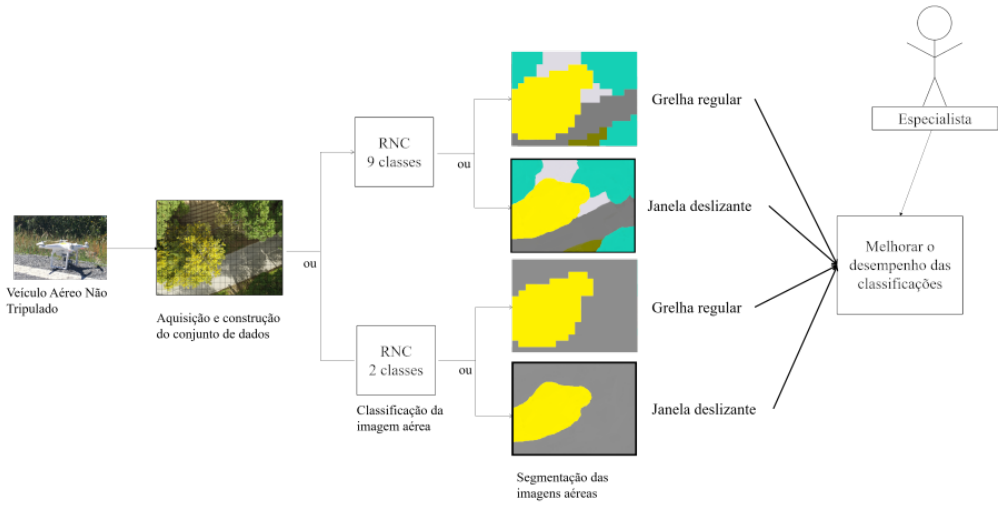



FIGURA 4.1: Arquitetura global da abordagem desenvolvida para o reconhecimento da espécie invasora *Acacia Longifolia*. O veículo aéreo captura imagens de alta resolução que são repartidas em sub-conjuntos de 200 x 200 píxeis para se realizar a classificação ou binária ou de múltiplas classes, obtendo-se uma segmentação da imagem aérea de alta-resolução. Os resultados da classificação podem ser melhorados através do recurso a um especialista para rever as previsões.

vez para se obter resultados mais detalhados. Na última etapa, o objetivo consiste em melhorar o desempenho das previsões realizadas pelo bloco anterior através da chamada a um especialista para corrigir as previsões cuja confiança é inferior a um limiar calculado para optimizar o compromisso entre a vantagem de melhorar a classificação das imagens obtidas e o custo de recorrer a um especialista para o efeito. Nas próximas secções serão descritas detalhadamente cada uma destas etapas.

## 4.1 Captura de imagens e conjunto de dados

Nesta secção são descritas as etapas para a construção do conjunto de dados fornecido [2] usado no contexto desta dissertação e ilustradas na Figura 4.2. A primeira etapa consiste na captura de imagens aéreas usando um veículo aéreo não tripulado, *DJI Phantom 3 Advanced quadcopter*, semelhante ao que se pode observar

[2]Esta etapa foi realizada pela empresa IntRoSys S.A..

na Figura 4.3. O planeamento dos trajetos a percorrer autonomamente pelo veículo aéreo foi realizado recorrendo ao software DroneDeploy. O plano do voo é posteriormente recebido pelo software de navegação executado a bordo do VANT.

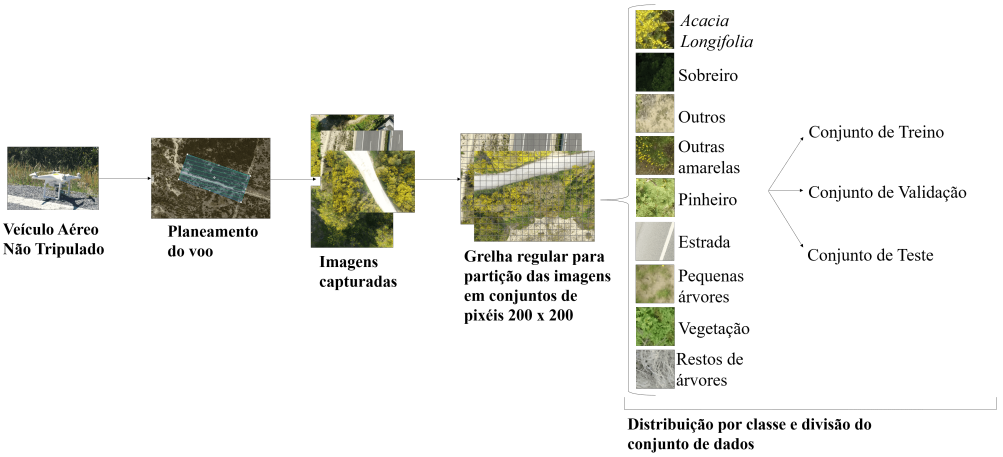
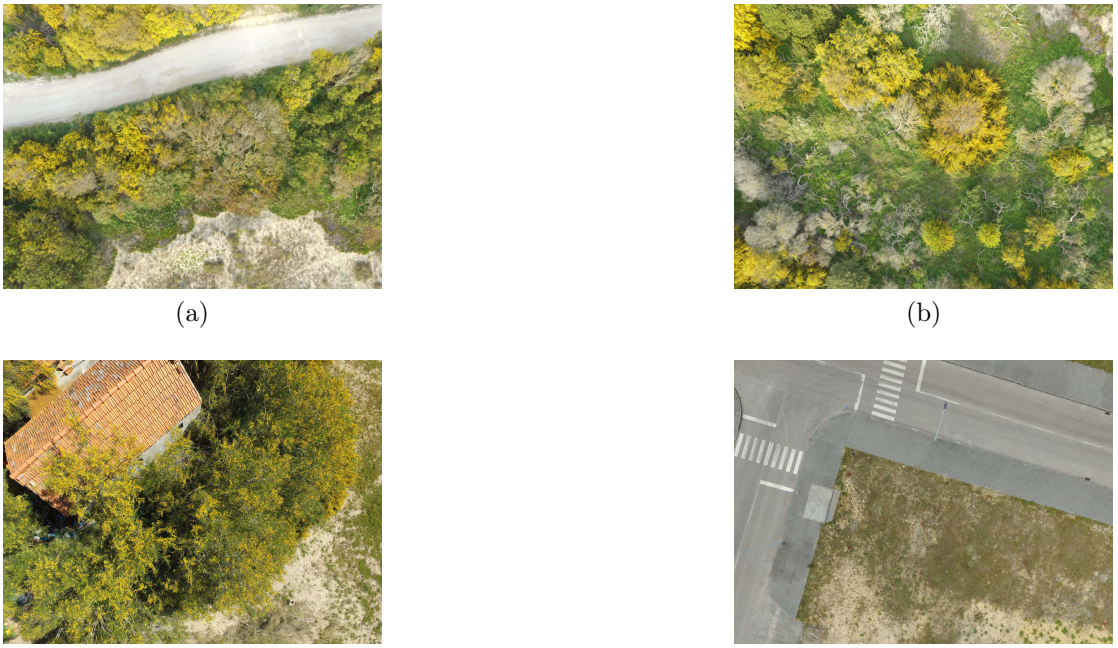



FIGURA 4.2: Etapas para a construção do conjunto de dados.

Os voos foram realizados no ano de 2016 em três localizações portuguesas diferentes: Costa da Caparica, Palmela e Sintra. O VANT percorreu 12 km, cobrindo uma área de quatro hectares. Durante os voos, as imagens foram adquiridas com uma câmara de 2.7 k a bordo, montada de forma a garantir que estivesse sempre apontada para baixo. As imagens foram adquiridas e armazenadas com uma resolução de 4000 x 3000 píxeis (ver exemplo de imagem adquirida pelo VANT na Figura 4.4).

FIGURA 4.3: Exemplo do veículo aéreo usado na validação experimental. Imagem retirada da página web [footage, 2019].

O conjunto de dados usado para o treino e validação do algoritmo é composto por 31 454 amostras. Cada amostra corresponde a uma sub-imagem de 200 x 200



(a) (b)

(c) (d)

FIGURA 4.4: Imagens aéreas capturadas pelo VANT.

píxeis obtidas através de uma grelha regular sobreposta a uma imagem semelhante à da Figura 4.4. Estas amostras foram classificadas manualmente consoante uma das nove classes (ver Tabela 4.1): Acacia L (7506 amostras), não Acacia L, nomeadamente, outras amarelas (1200 amostras), sobreiro (2912 amostras), pequenas ervas (3998 amostras), restos de árvores (1200 amostras), pinheiro (2922 amostras), vegetação (6238 amostras), outros componentes (2917 amostras) e estrada (2561 amostras). Por meio de uma amostragem aleatória, para uma segunda verificação, foi estimado um erro de 1.6% de classes incorretamente atribuídas durante a fase de construção do conjunto de dados [3].

Posteriormente, dividiu-se o conjunto de dados aleatoriamente em três sub-conjuntos: o conjunto de treino com 60% das amostras, o de validação com 20% das amostras, e de teste com os restantes 20% de amostras (ver Tabela 4.2). Na Figura 4.5 ilustram-se algumas amostras do conjunto de dados pertencentes a cada classe, sendo que especie *Acacia Longifolia* foi capturada em época de floração da espécie.

[3]Esta estimativa foi realizada pela empresa IntRoSys S.A.

| Classes | Número de amostras |
|---|---|
| Acacia L. | 7506 |
| Outras amarelas | 1200 |
| Sobreiro | 2912 |
| Pequenas ervas | 3998 |
| Restos de árvores | 1200 |
| Pinheiro | 2922 |
| Outros componentes | 2917 |
| Estradas | 2561 |
| Vegetação | 6238 |

TABELA 4.1: Número de amostras por classe.

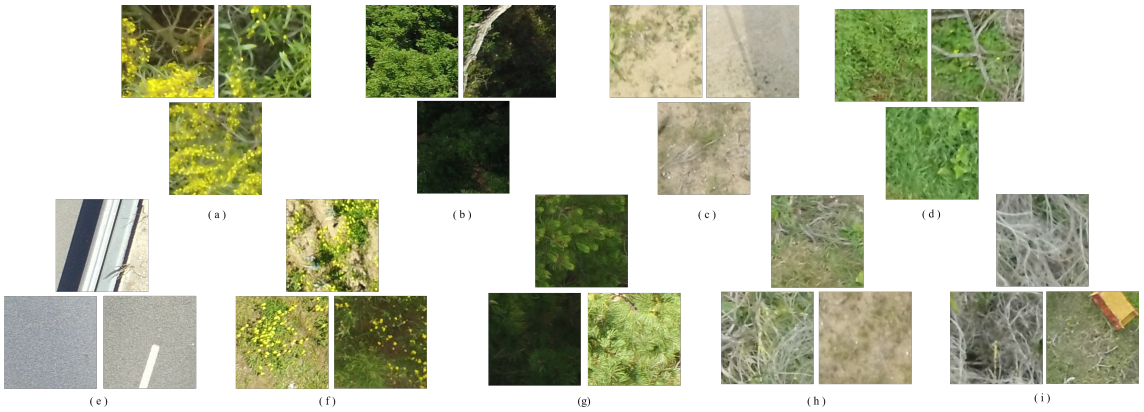



FIGURA 4.5: Conjunto de amostras pertencentes a cada classe. (a) Acacia L.; (b) Sobreiro; (c) Outros componentes; (d) Pequenas ervas; (e) Estrada; (f) Outras amarelas; (g) Pinheiros; (h) Vegetação e (i) Restos de árvores.

| Número de amostras | Conjunto de dados |
|---|---|
| 18872 | Treino |
| 6291 | Validação |
| 6291 | Teste |

TABELA 4.2: Distribuição de amostras nos conjuntos de treino, validação e teste.

## 4.2 Classificação automática da espécie *Acacia Longifolia*

Para o reconhecimento automático da espécie invasora *Acacia Longifolia*, foram desenvolvidas e treinadas duas redes neuronais convolucionais. Como referido no Capítulo 2, o desempenho das redes neuronais na classificação é influenciado pelos

valores dos hiperparâmetros. Deste modo, foram realizadas várias experiências variando-se os hiperparâmetros da rede até se alcançar a arquitetura final, com uma taxa de acerto e um desempenho em termos de tempo de computação desejados. Variou-se o número de camadas, a dimensão dos filtros de convolução, o número de épocas do treino, a probabilidade da desativação aleatória de neurónios, o algoritmo de optimização e o valor da taxa de aprendizagem. Na Tabela 4.3 são apresentadas as configurações dos ensaios realizados correspondentes à rede que realiza a classificação em nove classes, sendo que a configuração e arquitetura do ensaio com melhor taxa de acerto foi utilizada para a construção da rede que distingue duas classes.

| Experiência | Número de camadas convolucionais | Dimensão dos filtros | | Desativação aleatória dos neurónios | | Épocas | Algoritmo de optimização |
|---|---|---|---|---|---|---|---|
| | | 1ªcamada | Restantes camadas | Camadas escondidas | 1ª camada completamente convolucional | | |
| 1 | 4 | 3x3 | | 0.25 | 0.5 | 20 | RMSProp |
| 2 | | 5x5 | | | | | |
| 3 | 8 | 5x5 | 3x3 | | | 100 | |
| 4 | | 7x7 | 5x5 | | | | |
| 5 | 6 | | 3x3 | | | | |
| 6 | | | 5x5 | 0.2 | | | Adam |
| 7 | | | | | | | |
| 8 | | | | | | | |
| 9 | | | | | | | |
| 10 | | 11x11 | 5x5 | | | | |
| 11 | | 3x3 | | | | 200 | |
| 12 | | 7x7 | 5x5 | | | | |

| Experiência | Taxa de aprendizagem | Transformação aleatória de amostras | Tempo de treino | taxa de acerto | | Erro | | Sobre-ajuste |
|---|---|---|---|---|---|---|---|---|
| | | | | Treino | Validação | Treino | Validação | |
| 1 | $10^{-5}$ | Não | ≈27 minutos | 75.1% | 68.7% | 0.67 | 0.81 | Não |
| 2 | | | ≈28 minutos | 76.3% | 71.6% | 0.63 | 0.79 | Não |
| 3 | | | ≈1 horas | 80.8% | 70.9% | 0.52 | 0.85 | Sim |
| 4 | | | | 82.3% | 76.5% | 0.49 | 0.69 | |
| 5 | | | ≈3 horas | 87.2% | 80.0% | 0.36 | 0.57 | |
| 6 | | | | 89.5% | 86.6% | 0.29 | 0.37 | Não |
| 7 | | | | 89.5% | 87.3% | 0.29 | 0.35 | |
| 8 | | | | 91.7% | 88.9% | 0.23 | 0.30 | |
| 9 | $10^{-3}$ | | ≈2 horas | 97.5% | 90.4% | 0.07 | 0.5 | Sim |
| 10 | $10^{-5}$ | | | 91.3% | 89.7% | 0.24 | 0.27 | Não |
| 11 | | | | 93.8% | 87.2% | 0.17 | 0.39 | Sim |
| 12 | | Sim | ≈ 3 horas | 91.3% | 91.9% | 0.24 | 0.22 | Não |

Tabela 4.3: Configurações da arquitetura e treino do modelo das experiências realizadas.

A rede neuronal convolucional correspondente à primeira experiência era composta por quatro camadas convolucionais com todos os filtros de convolução de dimensões 3 x 3. O seu treino teve uma duração de 20 épocas utilizando o optimizador RMSprop com uma taxa de aprendizagem de $10^{-5}$. No entanto, esta configuração resultou numa taxa de acerto baixa e num erro bastante elevado. Por isso, a partir desta rede, realizaram-se alterações aos hiperparâmetros e verificou-se que

o aumento do número de camadas convolucionais implicava na ocorrência de sobre-ajuste aos dados. Esta conclusão é justificada pelo facto das experiências 3 e 4 apresentarem sobre-ajuste aos dados e o ensaio 6 não demonstrar este problema, sendo que a sua configuração é semelhante à configuração das outras experiências, distinguindo-se num número menor de camadas e apresentando um número de épocas de treino superior. O aumento da taxa de aprendizagem, introduzido na experiência 9, provocou um erro de generalização elevado, pelo que se manteve o valor de taxa de aprendizagem inicial. Observou-se pelos resultados dos ensaios 1 e 11 que um filtro composto por 3 x 3 píxeis influencia o aparecimento de sobre-ajuste aos dados justificado pelo facto de se detetar características específicas dos dados de treino que não permitem generalizar para amostras novas, devido à reduzida dimensão do filtro. Descartou-se a opção de recorrer a um filtro com dimensão 11 x 11 porque apesar de apresentarem uma taxa de acerto elevada, o seu tempo de treino para 100 épocas é bastante superior ao tempo nas restantes configurações dos filtros de convolução.

Com base na análise às experiências realizadas, concluiu-se que, de entre as várias configurações testadas, o modelo que melhor se adequa ao problema do reconhecimento da espécie invasora *Acacia Longifolia* corresponde a uma rede composta por seis camadas de convolução e três camadas de sub-amostragem através do valor máximo, com um filtro de sub-amostragem de 2 x 2 e deslocamento de 2 píxeis, sendo na verdade o tipicamente utilizado nas investigações com RNC [CS, 2019]. Cada uma destas camadas de sub-amostragem está intercalada por duas camadas convolucionais, sendo que a função de ativação escolhida para estas últimas camadas é a ReLu [Murphy, 2016]. No final da rede encontram-se duas camadas completamente conectadas para a previsão da classe da amostra de imagem recebida à entrada da rede. A primeira camada conectada usa a função de ativação ReLu, enquanto a segunda usa a função *softmax* para a classificação.

Foram desenvolvidas duas redes que diferem na última camada conectada para a realização de dois tipos de classificação: multi-classificação, onde a rede tem nove unidades na última camada conectada; e classificação binária, onde essa camada é composta por apenas duas unidades. Para reduzir o sobre-ajuste aos dados,

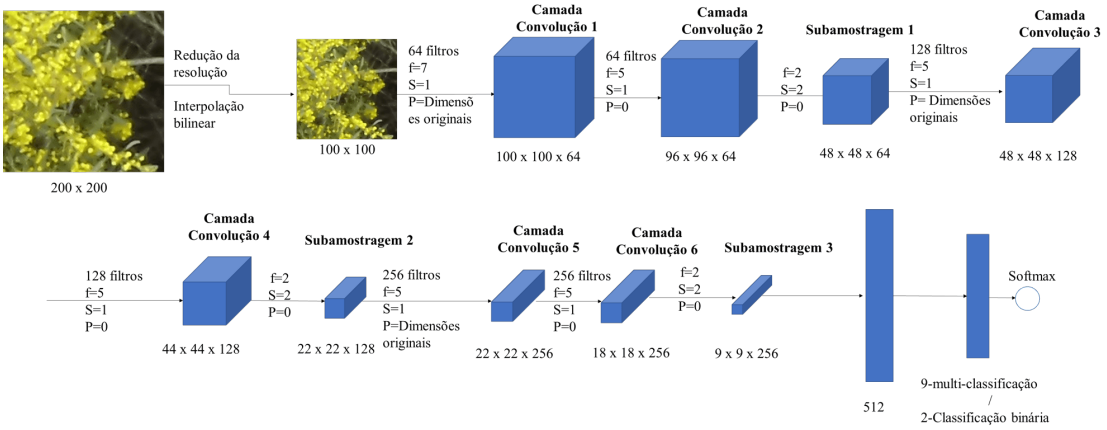



FIGURA 4.6: Arquitetura das redes neuronais convolucionais desenvolvidas para a classificação automática da espécie invasora *Acacia Longifolia*. A configuração da última camada altera-se consoante o tipo de classificação, apresentando nove unidades para a classificação múltipla e duas unidades para a binária.

incluíram-se camadas de desativação aleatória dos neurónios seguidas das camadas de sub-amostragem com valor de probabilidade de 0.2 e na primeira camada completamente conectada, com valor de 0.5. A arquitetura final da rede e as dimensões dos mapas de características resultantes, após as camadas de ativação, está retratada na Figura 4.6 e corresponde à experiência 12. A Tabela 4.4 resume a estrutura das duas configurações da arquitetura de cada rede. Para as camadas de convolução são apresentados três valores correspondentes ao número de filtros de convolução, à dimensão dos filtros e à função de ativação utilizada. Para as camadas de sub-amostragem é referida a dimensão do filtro de amostragem e a taxa de desativação de neurónios. Por último, nas camadas completamente conectadas informa-se o número de neurónios, a função de ativação e a taxa de desativação de neurónios da rede.

A camada de entrada, isto é, a primeira camada de convolução, recebe imagens RGB (três canais) com dimensão 100 x 100 píxeis para reduzir o tempo de computação. Como tal, antes de alimentar a rede, reduz-se a resolução das amostras do conjunto de dados fornecido com dimensão de 200 x 200 píxeis para 100 x 100 píxeis utilizando interpolação bilinear da biblioteca OpenCV. O número de filtros convolucionais na primeira camada é 64. Este valor é duplicado iterativamente em cada camada de convolução à medida que a rede se torna mais profunda e complexa. Na primeira camada, a dimensão do filtro é 7 x 7 e nas restantes camadas

| Camadas | CNNBin | CNNMulti |
|---|---|---|
| Camada Convolucional 1 | 64/7x7/ReLu | 64/7x7/ReLu |
| Camada Convolucional 2 | 64/5x5/ReLu | 64/5x5/ReLu |
| Camada de sub-amostragem máxima 1 | 2x2/0.2 | 2x2/0.2 |
| Camada Convolucional 3 | 128/5x5/ReLu | 128/5x5/ReLu |
| Camada Convolucional 4 | 128/5x5/ReLu | 128/5x5/ReLu |
| Camada de sub-amostragem máxima 2 | 2x2/0.2 | 2x2/0.2 |
| Camada Convolucional 5 | 256/5x5/ReLu | 256/5x5/ReLu |
| Camada Convolucional 6 | 256/5x5/ReLu | 256/5x5/ReLu |
| Camada de sub-amostragem máxima 3 | 2x2/0.2 | 2x2/0.2 |
| Camada completamente conectada 1 | 512/ReLu/0.5 | 512/ReLu/0.5 |
| Camada completamente conectada 2 | 2/*softmax*/0 | 9/*softmax*/0 |

TABELA 4.4: Configuração dos parâmetros das duas RNC. Camadas de convolução: número de filtros de convolução / dimensão dos filtros de convolução / função de ativação. Camadas de sub-amostragem: dimensão do filtro de sub-amostragem / probabilidade de desativar neurónios. Camadas completamente conectadas: número de unidades / função de ativação / probabilidade de desativar neurónios.

de convolução utiliza-se um filtro com menores dimensões de 5 x 5, seguindo-se a configuração comum de colocar o primeiro filtro com a maior dimensão [Patterson and Gibson, 2017]. São utilizados filtros de dimensões não muito elevadas nem reduzidas para se conseguir encontrar características específicas de cada classe, mas sem ocorrer o sobre-ajuste aos dados. Para além disso, como se observou pelas experiências já mencionadas, o aumento da dimensão dos filtros incrementa o tempo de treino.

Apesar do conjunto de dados se dividir em nove classes, o objetivo deste estudo é a deteção da espécie invasora Acacia L, ou seja, distinguir e reconhecer se uma dada amostra contém a espécie Acacia L. Como referido, foram treinadas duas redes neuronais convolucionais com a arquitetura referida anteriormente, diferenciando no número de unidades na última camada da rede, pois o tipo de classificação é diferente. Uma das RNC, designada por **CNNMulti**, é treinada para distinguir nove classes, recebendo por isso nove unidades na última camada. Este treino da rede trata-se de uma tarefa adicional à classificação binária para a distinção da espécie invasora *Acacia Longifolia*, sendo que a classificação da rede CNNMulti é transformada numa binária através da conversão da sua matriz de confusão de

nove classes numa de duas classes: Acacia L. e Não Acacia L. e pela qual se calcula o desempenho deste modelo para a classificação binária. O objetivo de treinar a rede CNNMulti consiste em perceber se a multi-classificação prejudica ou não o desempenho do modelo na deteção da classe Acacia L., para além disso, a classificação em multi-classes pode ser útil para o fornecimento de informação semântica sobre o ambiente, importante para controlar a evolução de espécies invasoras e os seus impactos na flora nativa. A segunda rede desenvolvida é treinada com a última camada conectada a receber dois neurónios para classificação binária, sendo designada por **CNNBin**. Através destas redes, estudou-se a influência da aprendizagem de multiclasses na tarefa principal do modelo, na deteção da espécie *Acacia Longifolia*, em comparação com os resultados obtidos pela rede binária.

### 4.2.1 Configuração do Treino

Antes de se realizar o treino das redes foi aplicado um pré-processamento aos dados com o objetivo de serem recebidos corretamente pelas redes. Começou-se por inverter a ordem das colunas e linhas da matriz recebida e composta por três matrizes a representar cada cor RGB de uma amostra. Esta matriz inicial, tinha uma dimensão de 3 x 100 x 100 (número de canais RGB=3, largura=100, altura=100) e converteu-se para a dimensão 100 x 100 x 3 (largura, altura, número de canais), respeitando a dimensão imposta pelo tensorflow. Assim a matriz resultante é composta por 100 sub-matrizes com 100 píxeis de linhas, sendo que cada linha apresenta três posições com informação de cada componente de cor RGB do pixel em questão, como se verifica na Figura 4.7. Os valores das matrizes anteriores foram normalizados para pertencerem ao intervalo entre zero a um, ajudando na diminuição do erro de generalização.

Em relação às classes categóricas originais, estas foram associadas a um número inteiro e guardadas num vetor. Na Tabela 4.5 encontram-se as classes categóricas associadas ao respetivo identificador do tipo inteiro. O vetor com a informação das classes é posteriormente convertido numa matriz composta por vetores organizados em linhas e cujo número de elementos é o mesmo que o número de classes. A

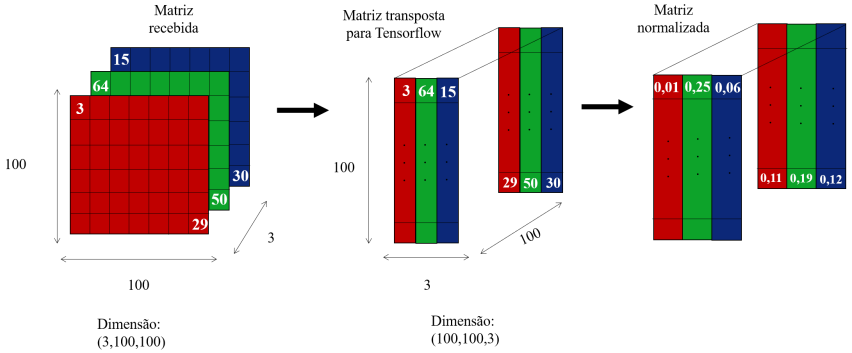



FIGURA 4.7: Pré-processamento dos dados recebidos antes do treino. Os valores 3, 64 e 15 são exemplos de valores RGB, respetivamente, que representam a cor de um pixel da amostra. Esses valores encontram-se separados em matrizes correspondentes a cada cor. As dimensões das matrizes são alteradas para as dimensões recebidas pelo tensorflow. Posteriormente, os valores são normalizados para facilitar o treino das RNC.

| **Classes categóricas** | **Classes numéricas** |
|---|---|
| Acacia L. | 0 |
| Sobreiro | 1 |
| Outros componentes | 2 |
| Outras amarelas | 3 |
| Pinheiro | 4 |
| Estrada | 5 |
| Pequenas ervas | 6 |
| Vegetação | 7 |
| Resto de árvores | 8 |

TABELA 4.5: Números inteiros associados às classes categóricas.

classe original é representada pelo valor um na posição associada ao valor que a representa, por exemplo, a classe Acacia L. está associada ao valor zero e por isso, o vetor de uma amostra desta classe apresenta na primeira posição o número um e valor zero nas restantes como se verifica na Figura 4.8. Na rede binária foi necessário converter as classes originais para binárias, ou seja, todas as amostras que não pertenciam à classe Acacia L. foram etiquetadas com Não Acacia L..

Nesta dissertação utilizou-se o algoritmo de taxa de aprendizagem adaptativa, Adam [Kingma and Ba, 2014] porque adapta a taxa de aprendizagem com base nos pesos e contribui no processo de aceleração da aprendizagem. A função de perda usada foi a entropia cruzada categórica [4] por se tratar de uma classificação

[4]Do inglês *categorical cross-entropy*

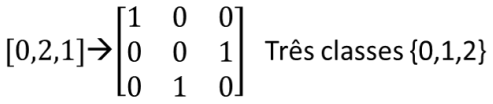



FIGURA 4.8: Exemplo da codificação das classes. Neste exemplo as classes originais encontram-se num vetor e de forma a serem lidas pelas RNC são transformadas numa matriz com número de linhas e colunas igual ao número de classes totais, em que cada classe verdadeira é representada pelo valor um na linha igual à posição da classe no vetor de classes e na coluna na posição igual ao valor da classe. Por exemplo, o primeiro valor do vetor de classes pertencente à classe 0 é representado na matriz através do preenchimento da primeira linha com valor um na posição 0.

de várias classes e porque se utilizou para o resultado da previsão a função de ativação *softmax*. O treino das redes foi realizado durante 200 épocas, porque se verificou que com estas configurações era possível treinar o algoritmo por mais tempo sem que este apresentasse sobre-ajuste aos dados.

O número de amostras por iteração [5] é uma potência de dois entre 32 e 256 [Goodfellow et al., 2016]. Se por iteração existir um grande conjunto de amostras, o poder da memória computacional exigido aumenta. Caso o conjunto tenha poucas amostras, existirá mais ruído e o tempo de treino será maior [Goodfellow et al., 2016]. Com estas informações utilizou-se o valor de 256 amostras por iteração, para evitar que o treino fosse lento, caso se selecionasse um valor demasiado pequeno de amostras por iteração. Com a utilização de um valor superior a 256, o treino do algoritmo seria impossível de ser executado devido às limitações associadas à capacidade de memória. A taxa de aprendizagem usada foi de $10^{-5}$, permitindo convergir sem apresentar um tempo de treino muito elevado, nem provocar o sobre-ajuste aos dados, como verificado na experiência 9.

Durante o treino da rede, aplicaram-se transformações às imagens em cada iteração, tais como, ampliação e redução da imagem de 50%, rotação entre zero a 45 graus e inversão horizontal e vertical. Os píxeis sem informação foram preenchidos por interpolação em espelho, para tornar as imagens mais naturais. Esta técnica de regularização permite reduzir a variância do modelo, ao serem aplicadas diferentes versões das imagens capturadas originalmente. Exemplos destas

[5] do inglês *batch size*

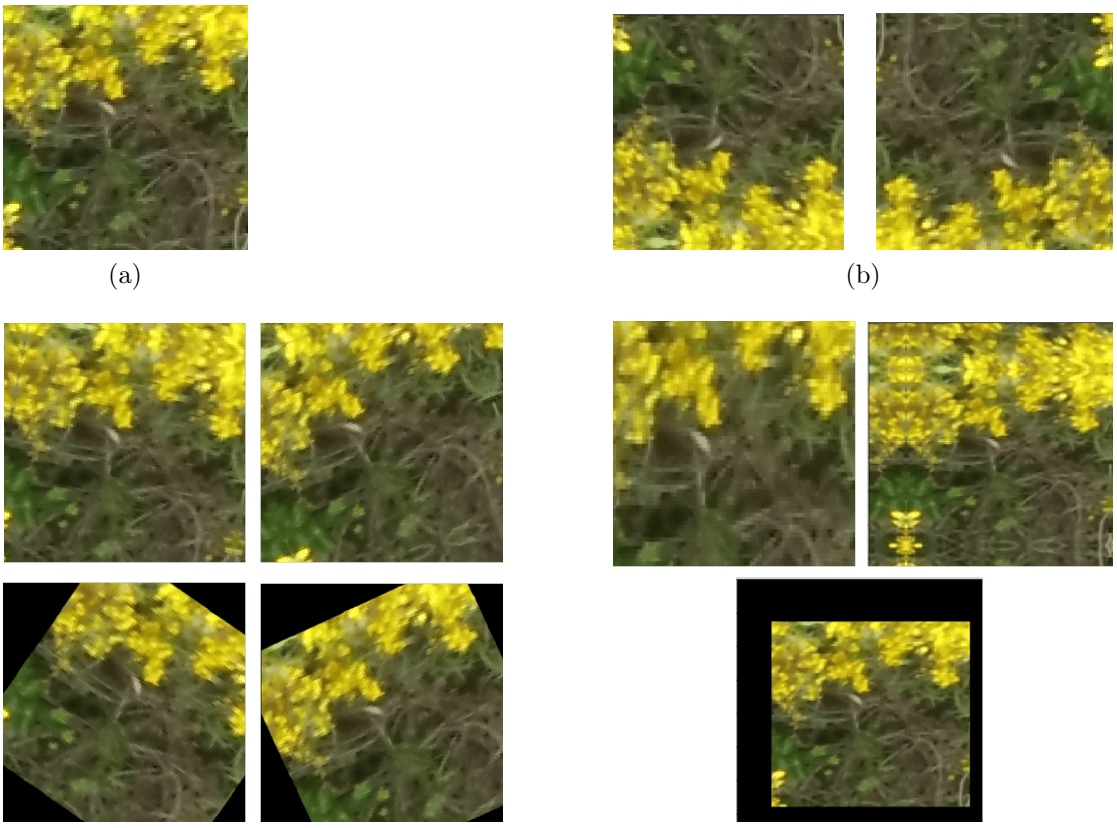


FIGURA 4.9: Exemplos de imagens geradas a partir de transformações da amostra original (a). (b): inversão horizontal e vertical, (c): rotação através do preenchimento dos píxeis (primeiras duas imagens de cima) e píxeis sem informação (últimas duas imagens) e (d): ampliação (direita), e redução da imagem com (em cima) e sem (em baixo) preenchimento dos píxeis sem informação.

transformações estão presentes na Figura 4.9. O treino, para cada uma das RNC proposta, demorou cerca de três horas.

### 4.2.2 Classificação das imagens obtidas pelo VANT

Após o treino e validação das RNC desenvolvidas, aplicaram-se os classificadores finais para realizarem a classificação e segmentação de imagens aéreas de alta resolução, com 4000 x 3000 píxeis, obtidas pelo VANT. Como as redes processam amostras de 200 por 200 píxeis, de forma a segmentar essas imagens de alta resolução é necessário desenvolver mecanismos adicionais. Existem duas abordagens para a resolução do problema.

**Grelha regular:** A técnica mais rápida consiste na amostragem da imagem de alta resolução recebida com uma quadrícula regular, representada através de uma janela deslizante, onde são extraídas amostras com dimensões de 200 x 200 píxeis, sendo que o deslocamento desta janela é também de 200 x 200 píxeis. Essas amostras são interpoladas para a dimensão de 100 x 100 píxeis a fim de serem recebidas pela rede. A classe prevista está associada a uma cor e todos os píxeis da amostra analisada serão representados por essa cor, como ilustrado na Figura 4.10. Com esta abordagem obtém-se uma imagem de baixa resolução com a segmentação da imagem original, sendo suficiente para os casos em que o VANT apenas necessita de uma estimativa pouco detalhada da presença da Acacia L..

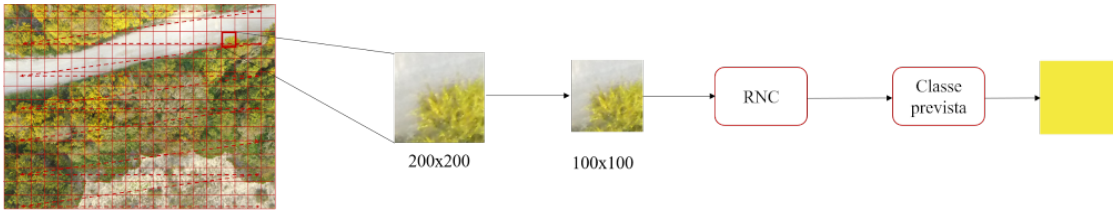



FIGURA 4.10: Ilustração da classificação através de uma grelha regular. Esta técnica consiste na divisão da imagem de alta resolução através de uma grelha regular e cada segmento de 200 x 200 píxeis é classificado pelas RNC. O segmento é pintado com a cor associada à classe. Neste caso a cor amarela representa a classe Acacia L..

**Classificação píxel a píxel:** Fazendo uso da técnica conhecida como janela deslizante com um deslocamento de um pixel de cada vez, a segmentação resultante apresenta-se mais detalhada, útil para estudar a localização e distribuição precisa das espécies invasoras. Neste caso, à medida que a janela deslizante percorre a imagem, o conjunto de píxeis cobertos pela janela são classificados pela RNC, ficando o pixel central da janela deslizante associado à classe prevista pela RNC e representado pela cor da respetiva classe. A janela de classificação desloca-se um pixel de cada vez antes de cada classificação, como ilustrado na Figura 4.11, e tem uma dimensão de 200 x 200 píxeis que é convertida para 100 x 100 antes de ser recebida pela RNC. O resultado ilustrado na Figura 4.11, é obtido através de várias classificações atribuídas a vários conjuntos de píxeis associados a janelas deslizantes.

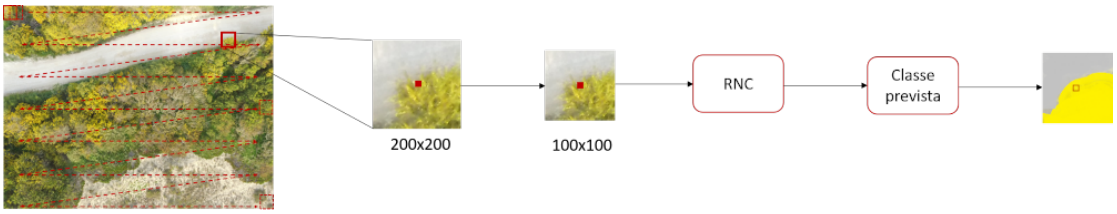



FIGURA 4.11: Ilustração da classificação através de uma janela deslizante com deslocamento de um pixel de cada vez. A janela deslizante percorre toda a imagem deslocando-se um pixel de cada vez e cada conjunto de píxeis abrangidos é classificado, representando somente o pixel central com a cor associada à classe prevista, como se ilustra pelo retângulo vermelho.

## 4.3 Aumento da taxa de acerto do modelo

O desempenho de uma rede neuronal convolucional pode ser melhorado através da comparação entre a curva da perda do treino e da validação, sendo a perda a diferença entre a classe verdadeira e a prevista, calculada de acordo com a Equação 2.1. Para se melhorar o desempenho da rede com base nessa análise, normalmente ajusta-se a configuração dos hiperparâmetros do algoritmo. Contudo, este método de exploração dos melhores valores para os hiperparâmetros implica re-treinar a rede a cada nova atualização, o que poderá provocar o sobre-ajuste aos dados de treino e de validação, para além de que o treino de uma rede neuronal convolucional apresenta um custo computacional elevado, resultando em fases de treino demoradas.

Nesta secção apresenta-se uma abordagem para aumentar o ganho da taxa de acerto das RNC na classificação de novos dados, sem se modificar os hiperparâmetros. Para tal, recorre-se a um mecanismo em que um especialista revê algumas das classificações previstas pela rede. Tendo em conta que as chamadas aos especialista acarretam um custo, é necessário limitar as classificações a serem verificadas.

Como as RNC podem devolver, no final da classificação, um vetor com as probabilidades de uma dada amostra pertencer às várias classes, através de um limiar de confiança [Zaragoza and d'Alché Buc, 1998], é possível controlar quando uma previsão deve ser revista pelo especialista, ou seja, quando a probabilidade da classe prevista é inferior ou igual a esse limiar considerado como aceitável.

Um limiar de confiança elevado resulta em corrigir mais classificações potencialmente erradas o que implica o aumento de chamadas ao especialista. Em contrapartida, um limiar com valores baixos acarreta um custo menor, pois existe um menor número de chamadas ao especialista para rever as previsões, contudo o valor da taxa de acerto terá um aumento menos significativo. O limiar de confiança com valores baixos é útil, por exemplo, para casos em que se aplica a classificação automática através de um VANT para o mapeamento de uma certa área, em que a relevância do ganho da taxa de acerto é prescindível e por isso, o número de chamadas ao especialista não necessita de ser elevado. No entanto, caso se queira garantir uma menor taxa de classificações incorretas, por exemplo, para eliminar certas plantas invasoras em campo, é necessário garantir que a espécie que se vai remover do ambiente é a correta. Neste caso, torna-se importante que as previsões sejam revistas por um especialista antes da ação, logo o limiar de confiança terá de ser elevado. Portanto, existe um compromisso entre a melhoria na taxa de acerto que resulta das chamadas ao especialista para a verificação das classificações previstas pela RNC e o custo associado a essas mesmas chamadas, seja ele temporal ou monetário.

Para gerir o compromisso referido no parágrafo anterior, o algoritmo calcula o limiar de confiança, $T_{min}$, que minimiza uma função de custo que pondera o benefício em termos de melhoria da taxa de acerto com o custo associado às necessárias chamadas ao especialista. Este limiar resulta de um processo de otimização para o qual o utilizador fornece um valor $\alpha \in [0, 1]$, correspondente à importância de chamar ou não o especialista. Desta forma, é possível associar mais ou menos relevância ao peso imputado ao número de chamadas ao especialista conforme o objetivo do utilizador.

Para um dado $\alpha$ definido pelo utilizador, o sistema encontra qual o limiar de confiança $T$ que oferece o melhor compromisso entre o ganho expectável na taxa de acerto em resultado de invocar o especialista, $G(T)$, e o custo associado a todas essas chamadas ao especialista. Desta forma, todas as amostras de confiança inferior ao limiar calculado pelo algoritmo $T_{min}$ serão revistas pelo especialista.

Formalmente, o ganho expectável na taxa de acerto é definido como:

$$G(T) = Acc(T) - AccO, \tag{4.1}$$

onde $Acc(T) = \frac{TP(T)+TN(T)}{D}$ representa a função que calcula a nova taxa de acerto para um dado limiar de confiança $T$, com $D$ a representar o número total de amostras do conjunto de treino e $TP$ e $TN$ como as amostras previstas corretamente e que dependem de $T$. AccO é a taxa de acerto original do modelo de classificação obtida com o conjunto de dados de treino, ou seja, antes de se receber qualquer ajuda por parte do especialista.

Para que o problema de encontrar o melhor compromisso possa ser definido como um problema de minimização de uma função de custo, é necessário reformular o ganho expectável de taxa de acerto como um custo, $C_A(T)$:

$$C_A(T) = -G(T). \tag{4.2}$$

O custo de chamar o especialista é definido como uma taxa entre o número de amostras revistas pelo especialista que depende do limiar de confiança $T$, $E(T)$, e o número total de amostras do conjunto de dados de treino, $D$:

$$C_C(T) = \frac{E(T)}{D}. \tag{4.3}$$

Através da Equação 4.2 relativa ao custo em melhorar a taxa de acerto e da Equação 4.3 do custo das chamadas ao especialista, formulou-se a função de custo que pondera a melhoria da taxa de acerto com a necessidade de chamar um especialista para rever as amostras, para um dado $\alpha$:

$$C(T, \alpha) = \alpha C_A(T) + (1 - \alpha)C_C(T). \tag{4.4}$$

Finalmente, o limiar de confiança $T_{min}$, para um dado $\alpha$ escolhido pelo utilizador, é obtido minimizando a função de custo $C(T, \alpha)$:

$$T_{min}(\alpha) = \underset{T \in \{0, 0.01, \ldots, 1\}}{\arg\min} \; C(T, \alpha). \tag{4.5}$$

Os limiares de confiança calculados pelo algoritmo estimam o valor de $T_{min}$ que minimiza o custo do compromisso para o conjunto de treino. De forma a perceber se o o limiar estimado durante a fase de treino $T_{min}$ é generalizável para outros conjuntos de dados da mesma distribuição para cada potencial $\alpha$, aplicou-se uma medida de erro para calcular a semelhança entre os limiares obtidos utilizando o conjunto de treino e o conjunto de validação. Esta medida de erro é calculada através do somatório da diferença quadrática dos limites associados a cada possível $\alpha$ entre os obtidos no conjunto de treino $T_{min}(\alpha)_t$ e no de validação $T_{min}(\alpha)_v$:

$$Erro = \sum (T_{min}(\alpha)_t - T_{min}(\alpha)_v)^2. \tag{4.6}$$

O algoritmo desenvolvido e referido nesta secção permite aumentar a taxa de acerto das classificações previstas pelo modelo baseado em redes neuronais convolucionais, consoante o valor de $\alpha$ escolhido pelo utilizador, que gere o compromisso entre o benefício de melhorar a taxa de acerto e o custo de realizar as necessárias chamadas ao especialista. O algoritmo desenvolvido calcula, para cada objetivo definido pelo utilizador, o limiar de confiança que limita o número de amostras que serão revistas pelo especialista. No próximo capítulo serão analisados os resultados desta investigação.

# Capítulo 5

# Resultados e Avaliação

Neste capítulo encontram-se apresentados os resultados do treino e validação das duas redes neuronais convolucionais apresentadas nesta dissertação. De forma a estudar o desempenho das redes, analisou-se a informação sobre a variação da função da perda entre a classe prevista pela rede e a sua verdadeira etiqueta e a evolução da taxa das previsões corretas ao longo de 200 épocas. A variação do erro na previsão das amostras de treino e de validação permite verificar a existência, ou não, de sobre-ajuste aos dados.

**Métricas de avaliação:** Para além da análise da evolução do treino das redes e da taxa de acerto no final do treino e nos dados de validação, gerou-se, no final do treino de cada rede, uma matriz de confusão com base nas classes previstas pelo modelo em questão. Esta matriz de confusão resultante da aplicação da rede ao conjunto de dados de validação permite visualizar o desempenho da rede na classificação, fornecendo informação sobre o número de imagens classificadas corretamente e incorretamente, por cada classe. Através da sua diagonal consegue-se estimar a qualidade do algoritmo na classificação, sendo que uma diagonal robustamente preenchida retrata uma RNC com elevada taxa de acerto. Para além disso, pela matriz de confusão é possível calcular a taxa de acerto das redes, através de $\frac{VP+VN}{VP+VN+FP+FN}$, em que $VP$ significa o número de verdadeiros positivos, ou seja, o número de imagens da classe *Acacia Longifolia* que foram classificadas corretamente, $VN$ de verdadeiros negativos, que corresponde a todas as imagens

da classe Não Acacia L. classificadas como tal, FP corresponde ao número de amostras de Não Acacia L. consideradas como Acacia L. e por último, $FN$ são as previsões erradamente classificadas como Não Acacia L., ou seja, as imagens da espécie invasora *Acacia Longifolia* que não foram detetadas. A soma de $VP$, $VN$, $FP$ e $FN$ corresponde ao número total de amostras. Como o objetivo é o reconhecimento da espécie *Acacia Longifolia*, é importante obter a percentagem de imagens que pertencem a esta espécie e que foram previstas corretamente. Para calcular esta taxa, recorre-se à equação da sensibilidade [1], de acordo com:

$$Sensibilidade = \frac{VP}{VP + FN}. \tag{5.1}$$

## 5.1 RNC para a classificação binária - CNNBin

Nesta secção estão presentes os resultados do treino e previsão da rede binária para o reconhecimento da espécie invasora *Acacia Longifolia*. Através da análise da Figura 5.1, que regista a função de perda durante o treino da rede binária, verifica-se que a rede não apresenta um sobre-ajuste aos dados, visível pela ausência da curva em forma de U da perda durante a previsão das amostras de validação.

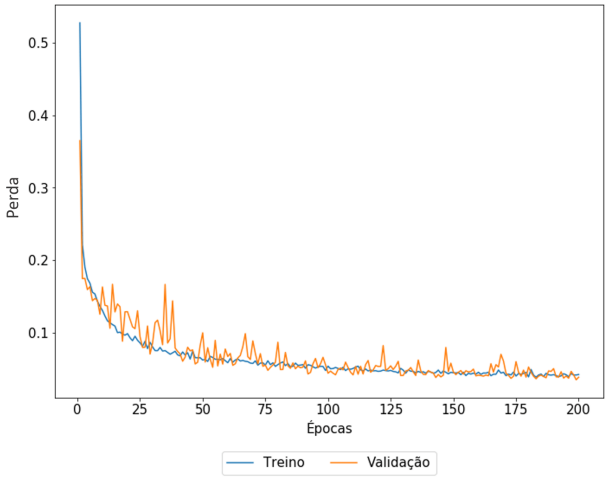



FIGURA 5.1: Evolução da perda ao longo do treino da CNNBin. As curvas do modelo mostram que este não entrou em sobre-ajuste aos dados, pois não existe uma diferença considerável entre o erro de treino e de validação, para além da ausência da curva em forma de U.

[1] Designado por *Recall*

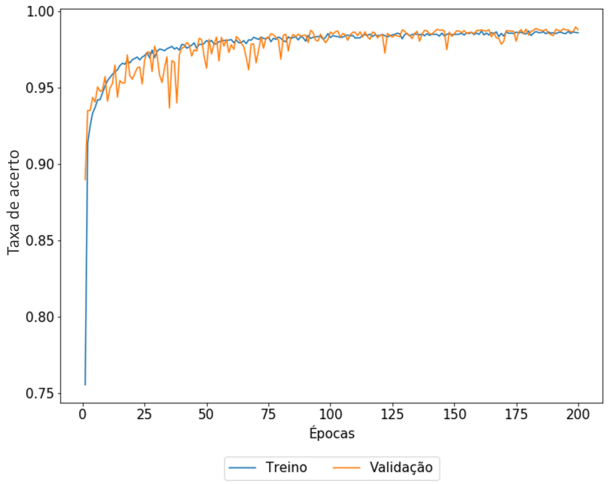



FIGURA 5.2: Evolução da taxa de acerto ao longo do treino da CNNBin. As curvas mostram o aumento da taxa de acerto ao longo do treino, tanto nos dados de treino como nos de validação evidênciando a inexistência de sobre-ajuste aos dados.

O valor da taxa de acerto na época 200, ou seja, no final do treino, é de 98.8% nos dados de validação, como se pode confirmar pelo número elevado de amostras corretas na diagonal da matriz de confusão, presente na Figura 5.3. A rede atinge uma sensibilidade de 95.5%, ou seja, reconhece 95.5% das imagens que pertencem à classe Acacia L. Estes resultados mostram a adequação das redes neuronais convolucionais para a deteção da espécie *Acacia Longifolia* em imagens aéreas capturadas por VANTs.

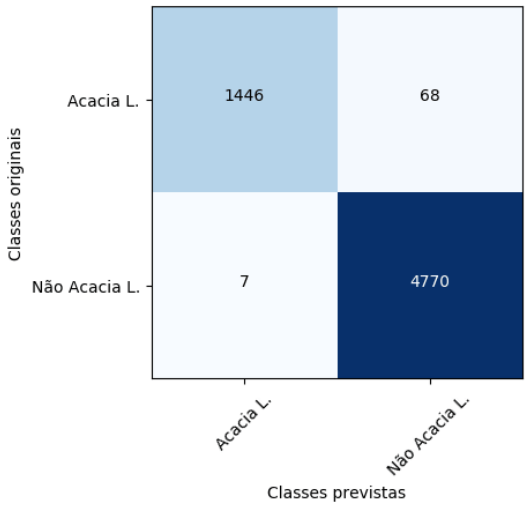



FIGURA 5.3: Matriz de confusão da rede CNNBin, utilizando os dados de validação.

## 5.2 RNC para a classificação múltipla - CNNMulti

Realizando a mesma análise que na rede CNNBin, a rede CNNMulti apresenta uma evolução da perda de validação decrescente, Figura 5.4. Tal como na rede anterior, não existe um aumento da perda face aos valores de treino, ou seja, a rede não se sobre-ajusta aos dados.

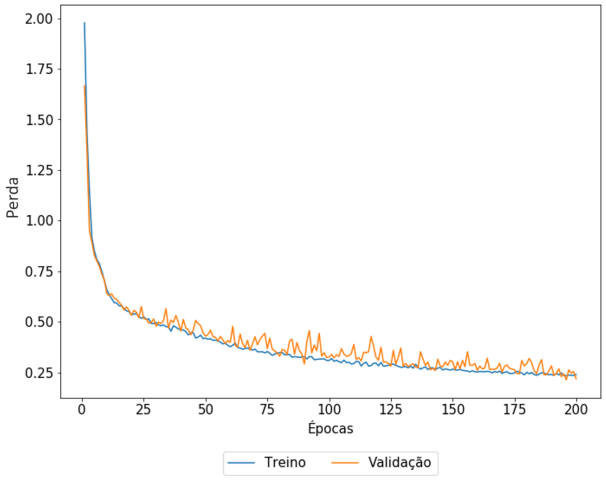



FIGURA 5.4: Evolução da perda ao longo do treino da CNNMulti. As curvas correspondentes ao erro de validação e treino demonstram a ausência de sobre-ajuste aos dados.

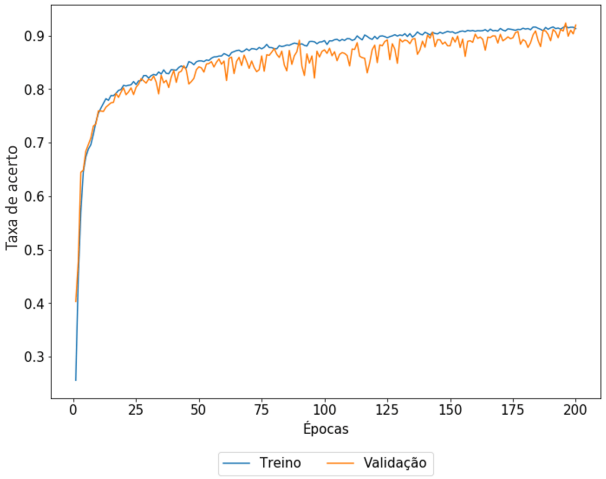



FIGURA 5.5: Evolução da taxa de acerto ao longo do treino da CNNMulti.

A taxa de acerto da rede no final do treino, nos dados de validação, é de 92%. Este elevado valor é confirmado pelo número de acertos em cada classe na matriz de confusão de multi-classificação apresentada na Figura 5.6. Estes resultados mostram a exatidão da rede em distinguir as nove classes.

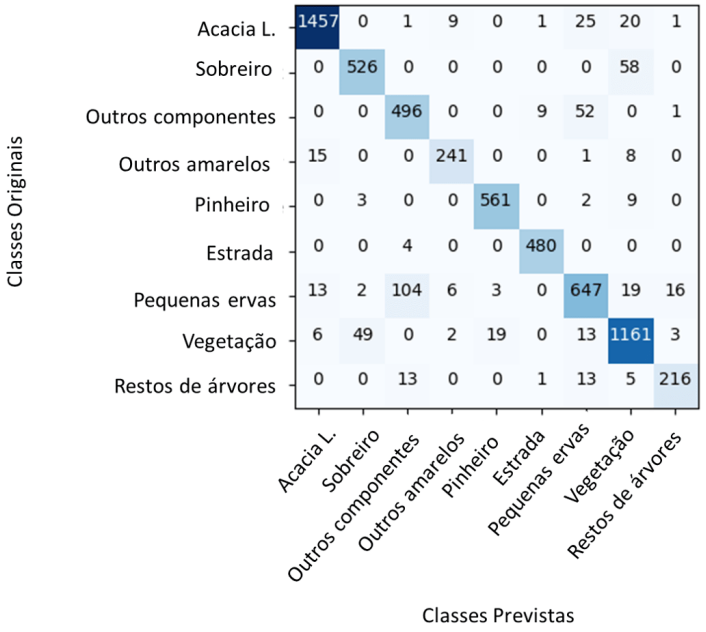



FIGURA 5.6: Matriz de confusão da rede CNNMulti utilizando os dados de validação.

Posteriormente, e de modo a analisar a influência da aprendizagem segundo nove classes na taxa de reconhecimento da espécie invasora em estudo, é fundamental calcular o verdadeiro valor da taxa de acerto da rede CNNMulti na tarefa principal, a classificação binária, já que a taxa de acerto fornecida após o treino, corresponde à classificação em nove classes. Para tal, convertem-se as etiquetas originais e as previstas de forma a representarem duas classes ao invés de nove e obtém-se a nova matriz de confusão para classificação binária, representada na Figura 5.7. Assim, possibilita-se a comparação entre os dois modelos desenvolvidos. Aplicando os dados presentes na matriz de confusão resultante, determinou-se a taxa de acerto, sendo que se obteve uma taxa de 98.6% de amostras classificadas corretamente e de 96.2% de amostras da espécie Acacia L. detetadas.

Os resultados obtidos mostram que a taxa de acerto entre a rede CNNBin e a CNNMulti difere de 0.2, sendo que a rede binária apresenta os valores superiores nas métricas de avaliação em comparação com a rede de multi-classes, podendo ser justificado pelo facto de se tratar da aprendizagem entre duas classes, tornando o processo de aprendizagem facilitado. Recorreu-se ao conjunto de amostras de

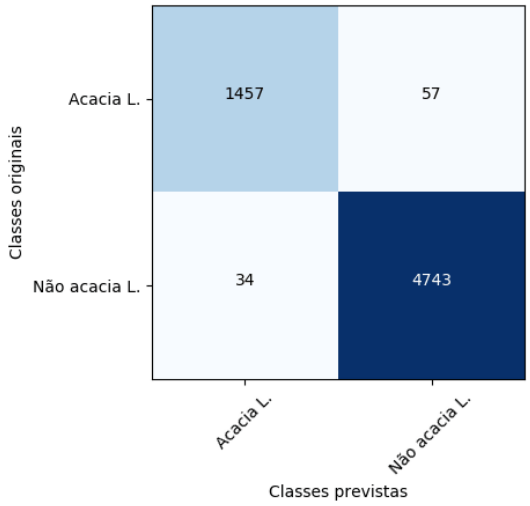



FIGURA 5.7: Matriz de confusão binária da rede CNNMulti.

teste para realizar uma última validação das redes. Como este conjunto de dados não foi utilizado nem para a computação do erro nem da taxa de acerto durante o treino, não existe probabilidade das redes se terem ajustado a esses dados. Os resultados completos obtidos utilizando o conjunto de dados de teste e a rede CNNBin podem ser verificados na Figura 5.8 e os resultados utilizando a rede CNNMulti na Figura 5.9b. A Tabela 5.1 sumariza os resultados das métricas de avaliação, tanto no conjunto de validação como no de treino. Como se pode verificar pela mesma, a taxa de acerto e de sensibilidade permanecem semelhantes. Com base nos resultados, conclui-se que a classificação segundo nove classes, não prejudica a rede na sua tarefa principal, ou seja, no reconhecimento da espécie invasora *Acacia Longifolia*.

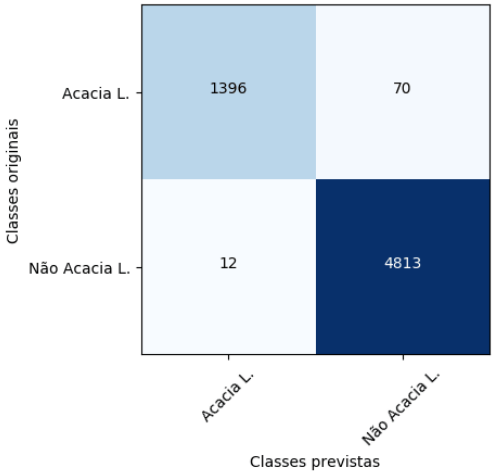
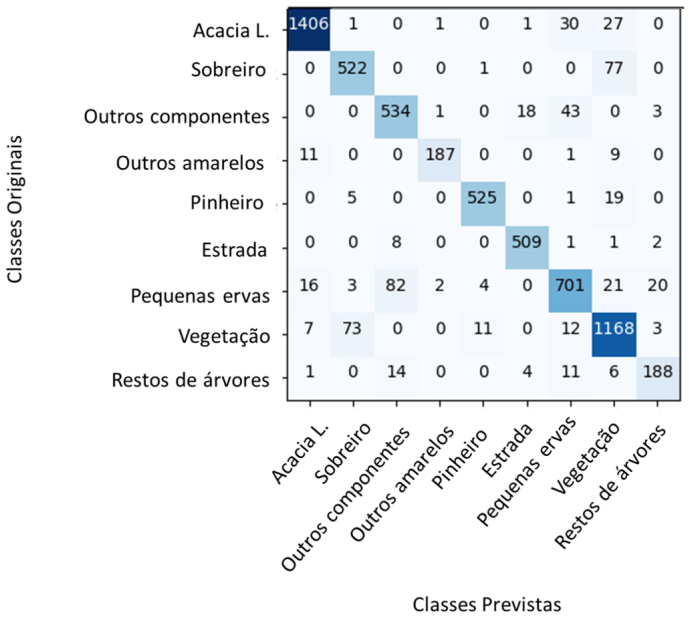
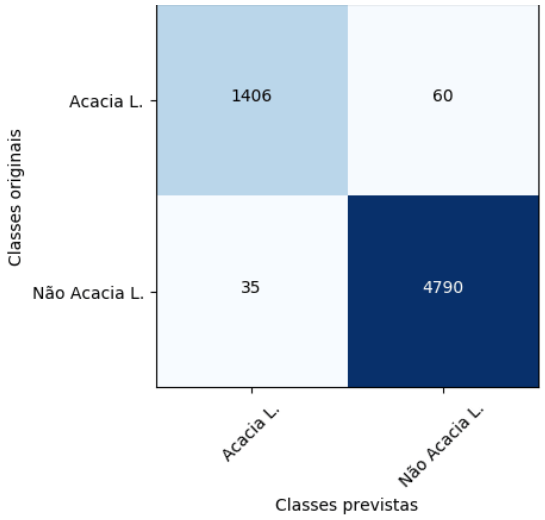



FIGURA 5.8: Matriz de confusão do conjunto de teste aplicado à rede CNNBin.




(a)




(b)

FIGURA 5.9: Matriz de confusão do conjunto de teste aplicado à rede CNNMulti: (a) classificação múltipla; (b) binária.

| | **CNNMulti** | | **CNNBin** | |
|---|---|---|---|---|
| | **taxa de acerto** | **Sensibilidade** | **taxa de acerto** | **Sensibilidade** |
| Conjunto de validação | 98.6% | 96.2% | 98.8% | 95.5% |
| Conjunto de teste | 98.5% | 96.0% | 98.7% | 95.2% |

TABELA 5.1: Sumário dos resultados da avaliação obtidos na rede binária e multi-classe usando os dados de validação e teste.

Na Figura 5.10 são apresentadas oito imagens selecionadas do conjunto de validação e outras oito amostras de teste. Existem quatro amostras, por conjunto de dados, classificadas incorretamente (cruz vermelha), duas corretamente (verificação verde). As amostras classificadas manualmente de forma incorreta durante a construção do conjunto de dados, mas classificadas corretamente pelas redes CNNBin e CNNMulti estão representadas pela letra L. Como se pode verificar, a Figura 5.10 apresenta alguns casos de previsões incorretas, explicadas pela presença de várias classes num só conjunto de píxeis 200 x 200. Nomeadamente, as imagens (c) e (g), que foram classificadas como Não Acacia L. devido ao facto da planta *Acacia Longifolia* se encontrar nos cantos da imagem. As imagens (d), (h), (l), e a imagem (p) são exemplos cuja sua classificação manual foi incorretamente atribuída como Acacia L., quando a sua verdadeira classe é pequenas ervas. Contudo, este erro demonstra a capacidade das redes em lidar com o ruído proveniente da classificação manual dos dados. Tanto a rede binária como a rede de multi-classificação conseguem classificar corretamente essas imagens, como pequenas ervas, na CNNMulti e Não Acacia L., na rede CNNBin.

Existem imagens com previsão incorreta, mas cuja probabilidade da classe prevista se aproxima do valor da probabilidade da classe correta, como se verifica na Tabela 5.2. Por exemplo, na amostra da Figura 5.10 (k), as probabilidades da classe Não Acacia L. e Acacia L. estão quase concentradas e igualmente distribuídas, próximas da probabilidade 50%.

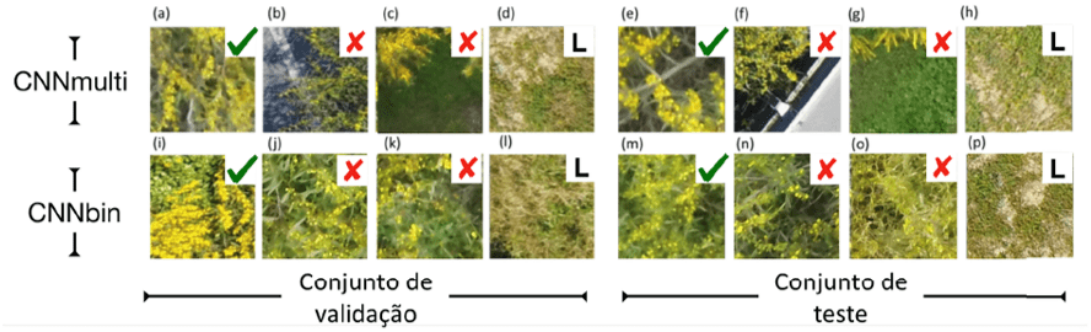
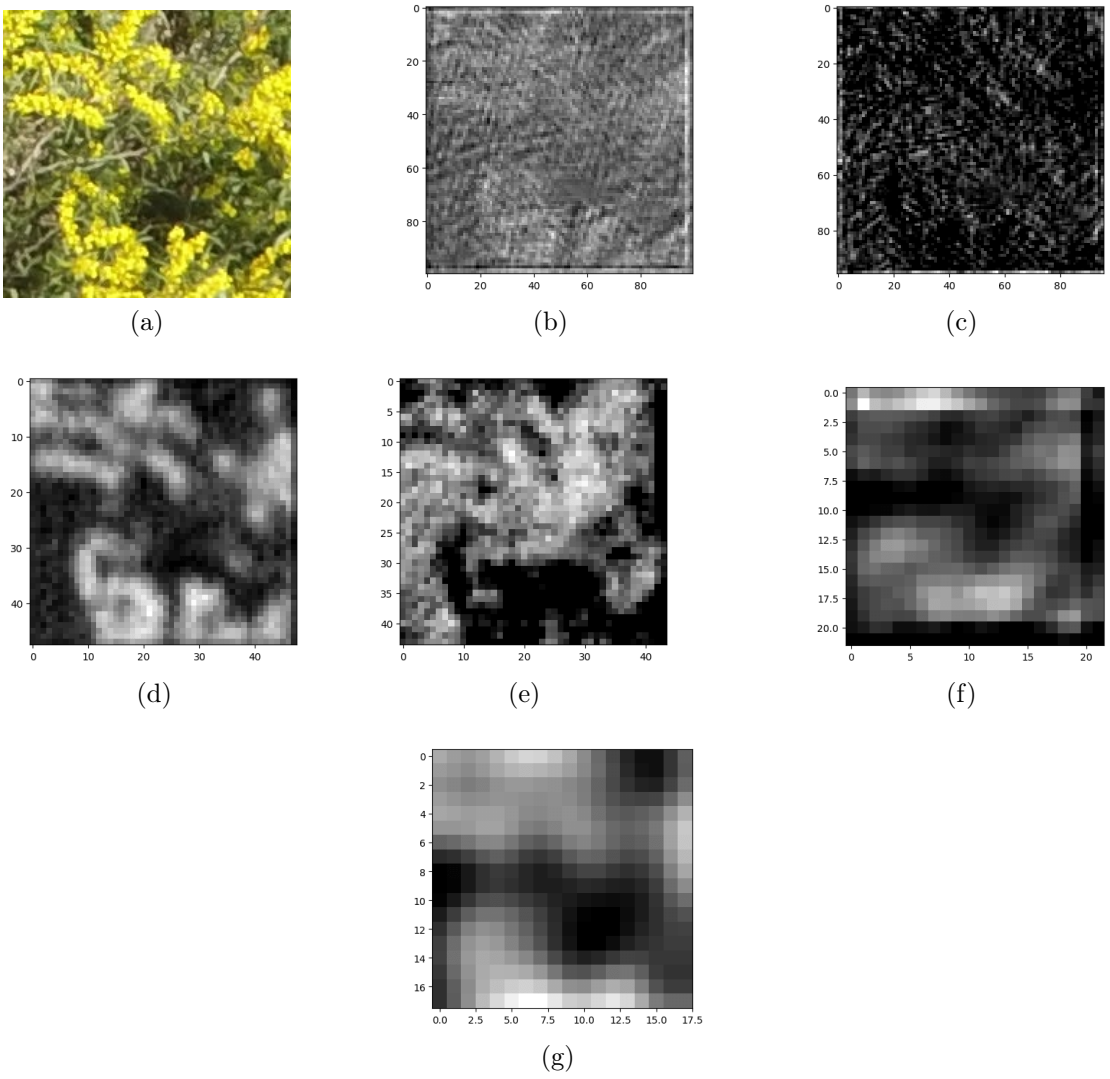



FIGURA 5.10: Amostras de imagens que foram classificadas pela rede CNNBin e CNNMulti.

| **Amostras** | **RNC** | **Probabilidades das classes** |
|---|---|---|
| Figura 7d | CNNMulti | [0: 2.739e-02; 1: 2.321e-06; 2: 7.942e-05; 3: 08.406e-05; 4: 8.467e-05; 5: 1.286e-6; **6: 0.969**; 7: 3.571e-03; 8: 7.412e-07] |
| Figura 7h | | [0: 9.971e-02; 1: 2.159e-06; 2: 6.552e-05; 3: 8.765e-04; 4: 2.661e-04; 5: 2.947e-06; **6: 0.892**; 7: 7.529e-03; 8: 5.019e-06] |
| Figura 7k | CNNBin | [**0: 0.486**; **1: 0.514**] |
| Figura 7l | | [0: 0.047; **1: 0.953**] |
| Figura 7o | | [**0: 0.461**; **1:0.539**] |
| Figura 7p | | [0: 0.0983; **1:0.902**] |

TABELA 5.2: Probabilidades das classes obtidas na previsão de algumas amostras, usando as duas RNC. Os números a negrito representam a classe prevista com maior probabilidade.

A estrutura da rede em camadas permite que uma rede neuronal aprenda características elaboradas a partir das obtidas nas camadas anteriores. À medida que se avança pela rede CNNMulti e CNNBin, os mapas de características apresentam as características que foram preservadas ao longo do treino e por isso, as suas dimensões são inferiores, comparativamente às dimensões dos mapas nas primeiras camadas de convolução devido à função de ativação e à sub-amostragem imposta que permite salientar as características importantes e gerais, como se ilustra nos píxeis a branco, pois são o que apresentam um valor de convolução maior, na Figura 5.11 relativa ao primeiro filtro de cada camada da rede CNNMulti.



(a) (b) (c) (d) (e) (f) (g)

FIGURA 5.11: Mapas de características da amostra original, aplicada à rede CNNMulti resultantes das várias camadas de ativação ao longo da rede de multi-classificação. (a) Amostra original recebida pela rede; (b) Resultado da primeira camada de convolução; (c) Resultado da segunda camada de convolução; (d) terceira camada de convolução; (e) quarta camada de convolução; (f) quinta camada de convolução; (g) sexta camada de convolução.

## 5.3 Classificação das imagens

A segmentação das imagens de alta resolução é realizada através de uma janela deslizante cujo conjuntos de píxeis abrangidos pela janela são classificados por uma das redes neuronais convolucionais desenvolvidas, dependendo do tipo de classificação. Cada classe prevista está associada a uma cor. Existem duas técnicas para a segmentação que diferem no valor de deslocamento da janela deslizante,

como referido na sub-secção 4.2.2 do Capítulo 4, obtendo-se segmentações com detalhes diferentes.

**Grelha regular:** A primeira técnica realizada através de uma grelha regular sobre a imagem original representa na Figura 5.12a, permite obter os resultados presentes nas Figuras 5.12b e 5.12c. Os resultados mostram uma segmentação menos detalhada sobre a área de interesse, como era de esperar pela aplicação de uma grelha regular, cuja janela deslizante avança 200 píxeis em cada deslocamento. No entanto, é útil para fornecer informação rápida sobre as espécies num dado local de monitorização.

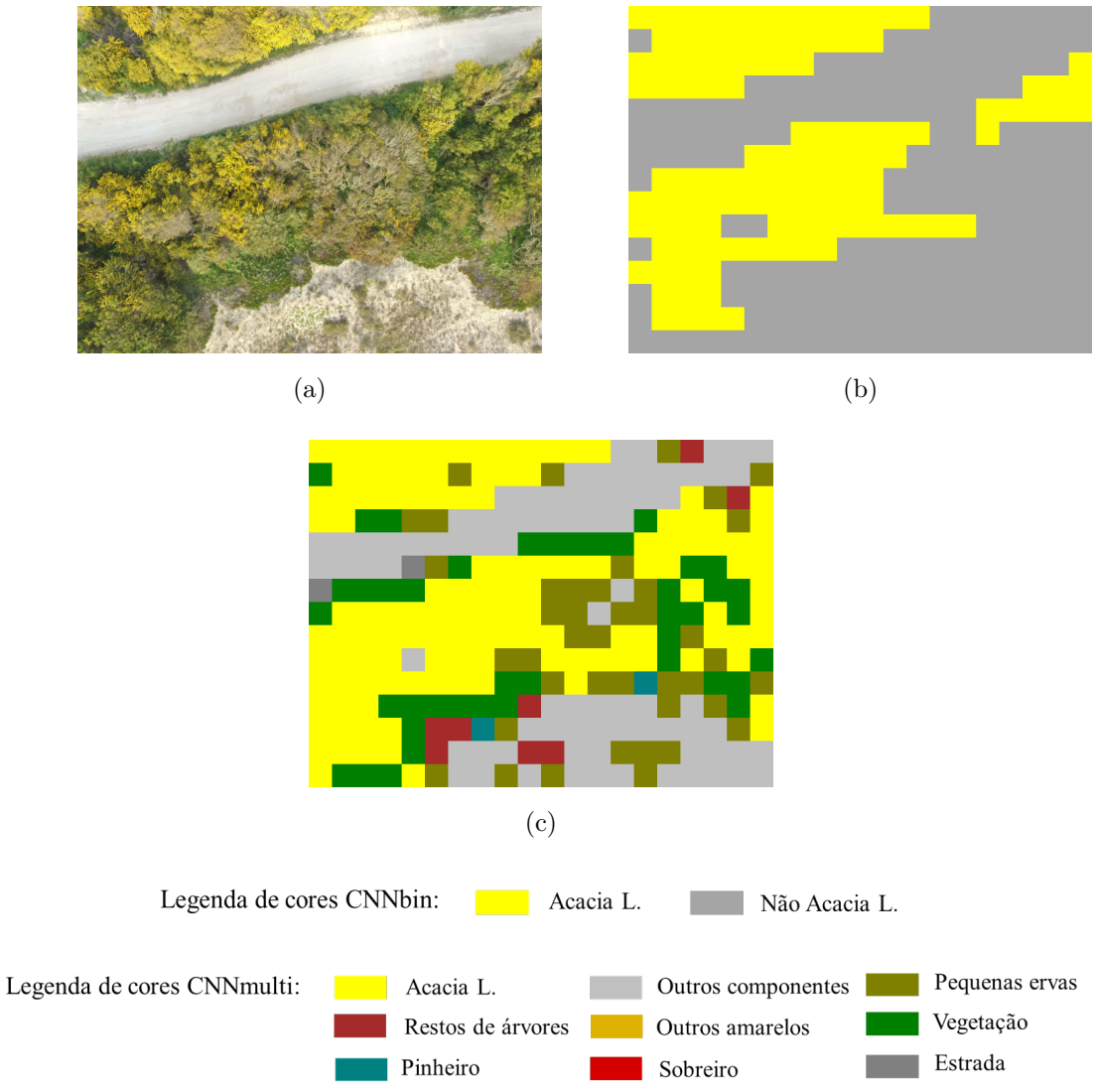



FIGURA 5.12: Segmentação através de uma grelha regular aplicada à imagem (a) usando a rede CNNBin (b) e a CNNMulti (c).

**Classificação píxel a píxel:** A segunda técnica designa-se por janela deslizante em que uma janela deslizante com dimensão 200 x 200 se desloca um pixel de cada vez obtendo-se uma segmentação com as diferentes classes bem definidas e sem ruído. Estes resultados mostram que o sistema não é muito sensível a pequenas variações provenientes da entrada, como se verifica pelas imagens 5.13b e 5.13c da Figura 5.13.

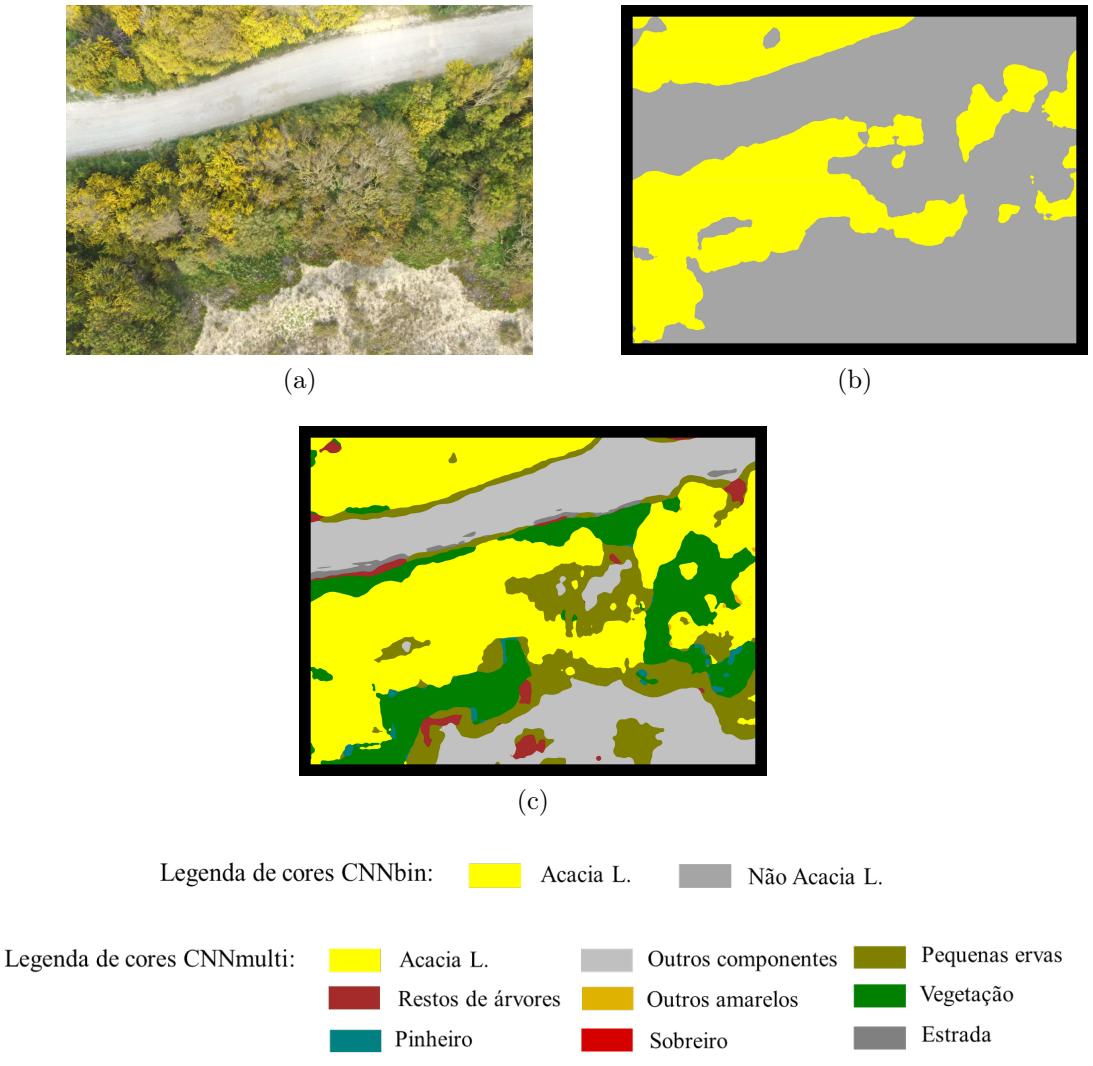



FIGURA 5.13: Segmentação através de uma janela deslizante, usando a rede CNNBin (b) e a CNNMulti (c).

## 5.4 Aumento da taxa de acerto do modelo

Nesta secção encontram-se os resultados do algoritmo desenvolvido que permite melhorar a taxa de acerto recorrendo a um especialista para verificar as amostras, sendo que se desconhece se a classificação prevista pela rede está correta ou não. Este algoritmo permite filtrar as amostras que serão ou não revistas, através do cálculo de um limiar de confiança, considerando as vantagens de uma possível correção de uma classificação prevista incorretamente com as desvantagens provenientes do custo de chamar um especialista para a classificação das previsões.

A importância de se verificar uma quantidade superior ou inferior de amostras previstas varia consoante a aplicação das redes neuronais convolucionais por parte do utilizador. Através de $\alpha$, que varia entre 0 e 1, o utilizador controla o peso do custo das chamadas ao especialista, sendo que quando $\alpha$ é zero, o utilizador está a dar máxima importância às chamadas ao especialista na equação que representa o custo do compromisso entre aumentar o número de previsões corretas e as chamadas ao especialista (verificar Equação 4.4). Como tal, o algoritmo associa ao valor de $\alpha$ escolhido pelo utilizador o limiar de confiança que minimiza o custo do compromisso em questão, otimizando assim, o número de amostras a serem revistas pelo especialista, através da Equação 4.5.

Para demonstrar o funcionamento do sistema no cálculo de $T_{mim}$ consoante o $\alpha$ escolhido pelo utilizador, simulou-se três valores de $\alpha$ e obteve-se para cada um, o respetivo gráfico com a evolução do custo do compromisso em função do limiar de confiança $T$, utilizando os dados de treino. Estes gráficos estão representados na Figura 5.14 e na Figura 5.15, para a rede CNNBin e CNNMulti, respetivamente, sendo que o valor do $T_{min}$ se encontra destacado por um ponto vermelho e corresponde ao limiar que permite o custo mínimo.

Após a aplicação do algoritmo nos dados de treino, verifica-se que na rede binária para $\alpha = 0.3$ o limiar de confiança que minimiza o custo é de 0.5, ou seja, todas as amostras cuja a probabilidade da classe prevista seja inferior ou igual a 50% serão revistas pelo especialista. Para $\alpha = 0.7$ e $\alpha = 0.9$ corresponde um limiar

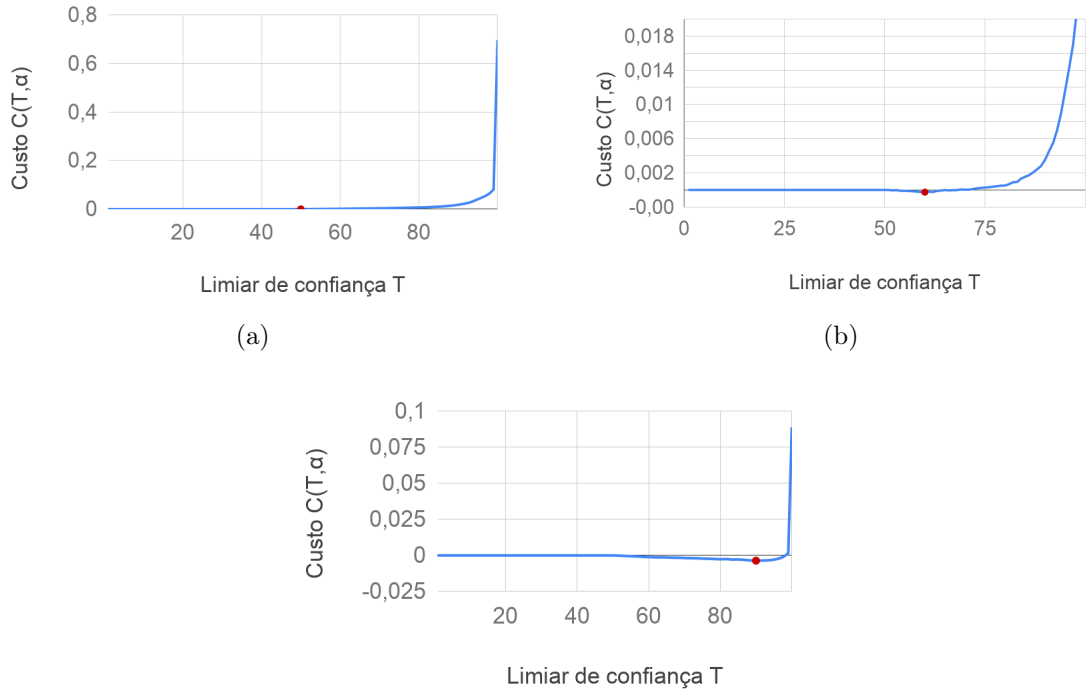



FIGURA 5.14: Custo calculado pelo algoritmo em função do limiar de confiança para $\alpha$ de 0.3 (a), 0.7 (b) e 0.9 (c) para a rede CNNBin.

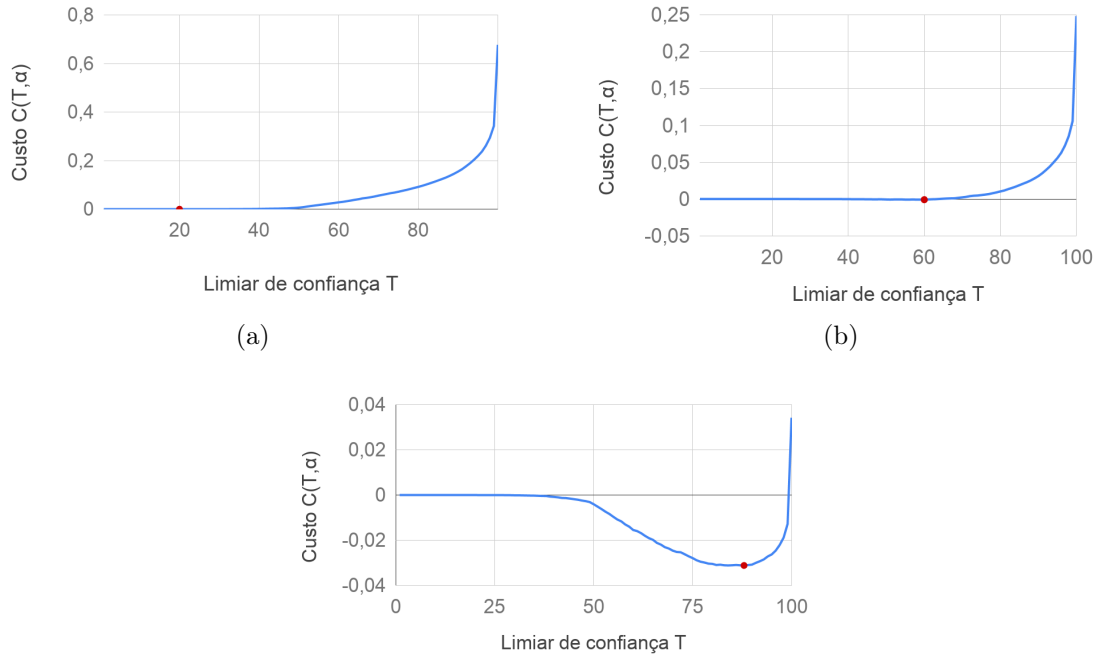
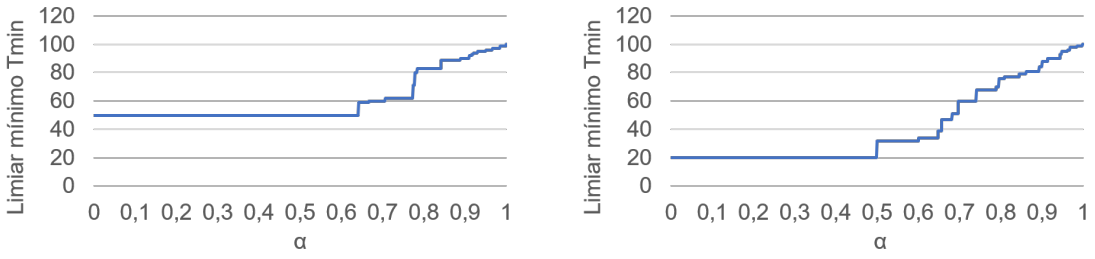



FIGURA 5.15: Custo calculado pelo algoritmo em função do limiar de confiança para $\alpha$ de 0.3 (a), 0.7 (b) e 0.9 (c) para a rede CNNMulti.

com probabilidade de 60% e de 90%, respetivamente. Em relação à rede de multi-classes, para um peso $\alpha = 0.3$, o limiar calculado pelo algoritmo é de 20%, enquanto que para um peso de 0.7 e 0.9 é de 60% e de 88%. Na rede que realiza a multi-classificação, o limiar de confiança para $\alpha = 0.3$ é inferior ao da rede binária porque a classe prevista poderá apresentar um valor de probabilidade inferior a 50%, nos casos em que o modelo apresenta maior dificuldade na classificação. Enquanto que na binária, para uma classe ser considerada a prevista para a amostra classificada, terá de ter uma probabilidade superior a 50%.

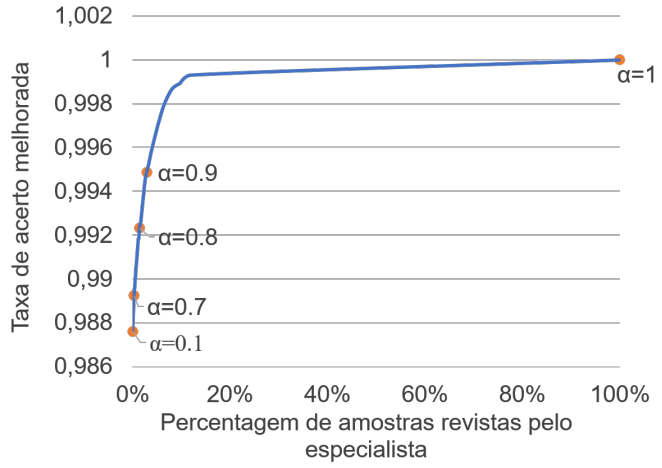
Desta forma, o limiar não é fixo, já que é calculado para um dado $\alpha$ escolhido pelo utilizador, de forma a se optimizar o valor do custo do compromisso final entre a possível correção de amostras classificadas incorretamente e as chamadas ao especialista, obtendo-se um limiar que limita quais as amostras que devem ser revistas. Na Figura 5.16 encontram-se os limiares de confiança que minimizam o custo para cada possível $\alpha$, calculados pelo algoritmo. Como se pode verificar, à medida que o peso do número de chamadas ao especialista diminui, ou seja, o $\alpha$ aumenta, o limiar de confiança que minimiza o custo do compromisso também aumenta. Este fenómeno é justificado pelo facto de que, como o $\alpha$ aumenta implica que o custo de melhorar a taxa de acerto tenha mais peso e para minimizar esse custo, o limiar de confiança tem de aumentar para garantir que a nova taxa de acerto após a aplicação do limiar de confiança seja cada vez mais superior à taxa de acerto inicial do sistema. Já quando o $\alpha$ favorece a parcela do custo da taxa de chamadas ao especialista, o limiar de confiança diminui de forma a garantir que quantidade de amostras a serem revistas seja pequena.



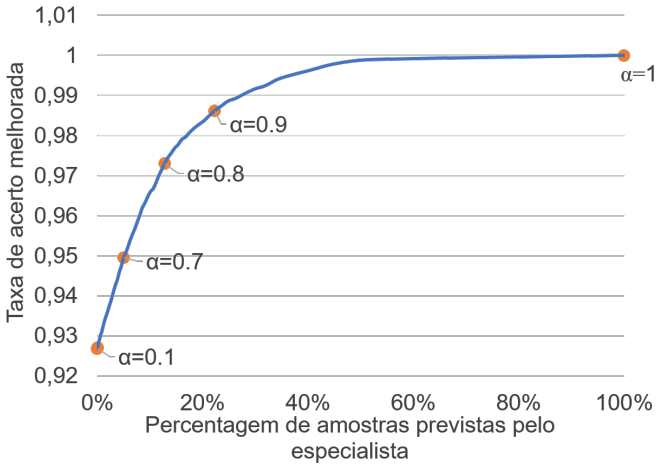

(a) (b)

FIGURA 5.16: Limite que minimiza o custo para cada $\alpha$ em função do $\alpha$ na rede CNNBin (a) e CNNMulti (b).

O algoritmo foi iterador para vários $\alpha$ de forma a se estudar a evolução da nova taxa de acerto em função do número de amostras revistas pelo especialista, relacionando com os vários pesos que podem ser definidos pelo utilizador $\alpha$. A Figura 5.17 representa o aumento da nova taxa de acerto associada à percentagem de amostras revistas no conjunto de treino de 18872 amostras, para os dois modelos de classificação desenvolvidos.




(a)




(b)

FIGURA 5.17: Melhoria da taxa de acerto em função do número de chamadas ao especialista na rede CNNBin (a) e CNNMulti (b).

Como se esperava, verifica-se um aumento abrupto da taxa de acerto seguido de uma estabilização, para os dois modelos de classificação automática, já que o aumento do número de amostras a serem revistas permite aumentar a nova taxa de acerto. No entanto, esta melhoria da taxa de acerto é abrupta até à verificação, por parte do especialista, de 20% de amostras do conjunto de dados de treino para

a rede CNNBin e a partir de 50% para a rede CNNMulti. Este acontecimento é explicado pelo facto de que a partir dos valores de percentagem de amostras referidos anteriormente, a quantidade de amostras classificadas incorretamente pelo modelo e recebidas para serem revistas, é bastante inferior comparativamente com as classificadas corretamente. Para além disso, o modelo binário apresenta uma elevada taxa de acerto, e por isso a curva do aumento da nova taxa de acerto surge com um crescimento mais acentuado do que na rede de multi-classificação. Como a taxa de acerto do sistema para a distinção das nove classes é bastante inferior, a curva da evolução da nova taxa de acerto para este modelo apresenta um aumento suave pois requer uma maior percentagem de amostras a serem revistas, cerca de 50% de amostras do conjunto de treino. Desta forma, compensa ao utilizador selecionar um $\alpha$ inferior a um, porque a taxa de acerto tem um aumento mais considerável nesses valores possíveis de $\alpha$, correspondendo a uma revisão por parte do especialista até 20% ou 50% das amostras do conjunto, caso se recorra ao modelo de multi-classificação ou binário, respetivamente.

O método aqui apresentado para melhorar a taxa de acerto apresenta-se previsível porque à medida que $\alpha$ aumenta a nova taxa de acerto também aumenta, permitindo que o utilizador consiga controlar o efeito da variação do $\alpha$, tendo conhecimento que incrementar o $\alpha$, para os dados de treino, implica sempre num aumento da taxa de acerto. Contudo, esta análise foi realizada utilizando o conjunto de dados de treino, pelo que é necessário verificar se os resultados obtidos pelo algoritmo desenvolvido podem ser generalizados para novos dados. Para tal, calculou-se o erro que fornece a diferença entre os limiares optimizados obtidos pelo algoritmo entre os dados de treino e de validação para cada possível $\alpha$, seguindo a Equação 4.6.

Os resultados do erro foram de 1.09899 e de 6.2399, para a rede CNNBin e CNN-Multi, respetivamente. O erro máximo possível é de 1001, pois a diferença entre os limiares de confiança que minimizam o custo nos dois conjuntos de dados é máxima quando os limiares são extremos um do outro, ou seja, quando um é zero e o outro um, o que implica numa diferença ao quadrado de um a múltiplicar pelos 1001 valores de $\alpha$ disponíveis. Tendo em conta este erro máximo, o erro calculado nas

duas redes é ínfimo correspondendo a 0.0011% e 0.0062% do erro total, para a rede binária e de múltipla classificação. Estes valores sugerem que o limiar resultante do conjunto de treino pode ser generalizado para os dados de validação e posteriormente, em dados não observados durante essa fase. Como era de esperar, o erro apresenta um valor superior na rede CNNMulti, apresentando maior dificuldade em determinar um limiar que generalize adequadamente, face à binária.

Com base nos resultados e nas observações realizadas, para se aumentar o valor da taxa de acerto é necessário recorrer a um especialista para rever e corrigir as amostras, para tal, esta solução acarreta um custo cujo peso da taxa de chamadas ao especialista varia consoante o valor de $\alpha$ escolhido pelo utilizador. Desenvolveu-se um algoritmo que dado um $\alpha$, fornecido pelo utilizador, devolve o limiar de confiança que permite minimizar o custo do compromisso em melhor a taxa de acerto da classificação e o custo de chamar o especialista. É este limiar de confiança que permite melhorar o desempenho final das previsões ao limitar as classificações a serem revistas pelo especialista, sendo que quanto maior for o valor desse limiar, melhor será o resultado da métrica taxa de acerto pois existem mais amostras a serem corrigidas, com a consequência do aumento do custo das chamadas ao especialista.

# Capítulo 6

# Conclusões

Esta dissertação teve como objetivo a construção e estudo de um sistema capaz de reconhecer automaticamente a espécie invasora da flora, *Acacia Longifolia*, em imagens aéreas capturadas por veículos aéreos não tripulados. A arquitetura do sistema que foi desenvolvida consiste na captura das imagens aéreas através de veículos aéreos não tripulados, no pré-processamento dessas imagens e na classificação automática de imagens aéreas para o reconhecimento da espécie invasora *Acacia Lonfolia*. Esta classificação baseia-se numa rede neuronal convolucional pois este tipo de classificadores não requer a seleção prévia de características pois esse processo é realizado pela própria rede ao longo do seu treino. As RNC apresentam esta vantagem em relação aos classificadores clássicos da aprendizagem automática, e para além disso, permitem alcançar valores de precisão e de taxa de acerto tipicamente superiores. No entanto, requerem para seu treino um extenso conjunto de dados previamente etiquetados.

Foram desenvolvidas duas RNC que diferem no número de classes a prever, uma foi treinada para classificação binária (CNNBin) entre Acacia L. e Não Acacia L. e outra para a distinção de nove classes (CNNMulti): Acacia L., Outras amarelas, Sobreiro, Pequenas ervas, Restos de árvores, Pinheiro, Vegetação, Outros componentes e Estrada. Posteriormente, a classificação multi-classe foi convertida para duas classes: Acacia L. e Não Acacia L. Com base na sua matriz de confusão

calcularam-se as métricas de avaliação para avaliar o desempenho da rede na função principal. Para se responder à questão de investigação sobre a topologia da RNC que melhor se ajusta ao problema, foram realizados estudos usando vários valores dos hiperparâmetros de configuração da arquitetura e do treino do algoritmo, comparando-se os resultados para atingir uma taxa de acerto elevada. De entre as várias configurações testadas, a que apresentou a melhor taxa de acerto corresponde a uma rede com seis camadas convolucionais, com 64 filtros de dimensão 7 x 7 píxeis na primeira camada e 5 x 5 nas restantes, e com duas camadas completamente convolucionais. O tempo de treino foi de 200 épocas, com uma taxa de aprendizagem de $10^{-5}$ e usando o ADAM como algoritmo de optimização. Para além disso, utilizaram-se como técnicas de regularização a desativação aleatória de neurónios e a geração de amostras sintéticas baseadas na aplicação de transformações das imagens originais. Utilizando estas técnicas a rede não apresentou sobre-ajuste aos dados durante o treino e validação, não sendo por isso necessária a aplicação de métodos de regularização adicionais, respondendo assim, à outra questão de investigação sobre a regularização necessária para o algoritmo de classificação automática desenvolvido.

Em relação à taxa de previsões de imagens pertencentes à classe *Acacia Longifolia* classificadas corretamente, ou seja, em relação à métrica sensibilidade, a rede binária detetou corretamente 95.2% das amostras de teste pertencentes à espécie invasora em estudo e a rede de múltipla classificação reconheceu 96.0%. Estes resultados demonstram que o uso de um classificador para a deteção de várias classes não deteriora o desempenho do sistema na tarefa da classificação binária primária, obtendo-se um valor de taxa de acerto utilizando o conjunto de dados de teste de 98.5%, diferindo duas décimas da taxa de acerto da binária, de 98.7%. Este elevado valor de amostras classificadas corretamente demonstra a validade da abordagem e, consequentemente, a viabilidade do uso de veículos aéreos para o mapeamento automatizado das espécies invasoras consideradas. Deste modo, através de um veículo aéreo e o classificador de várias classes pode-se obter adicionalmente uma descrição semântica do ambiente sem prejudicar a capacidade de detetar com

precisão a flora pertencente à espécie *Acacia Longifolia*, respondendo-se assim à penúltima questão de investigação.

Por último, relativamente à questão do mecanismo que permite aumentar a taxa de acerto, foi desenvolvido um método que filtra as previsões que devem ser verificadas por um especialista através do cálculo do valor de limiar de confiança que optimiza o custo do compromisso entre melhorar as classificações e o custo das chamadas ao especialista. Desta forma, a escolha do utilizador está associado a um limiar de confiança optimizado a partir do qual as previsões abaixo desse valor serão revistas pelo especialista e consequentemente o valor da taxa de acerto aumenta.

Com base nos resultados e tendo em conta as investigações presentes no Capítulo do Estado da Arte sobre a aplicação de RNC para o reconhecimento de espécies invasoras em imagens aéreas, verificou-se que a maioria das investigações recorrem à aplicação de algoritmos de aprendizagem automática clássicos. Poder-se-á concluir que a utilização de rede neuronais convolucionais, nomeadamente as desenvolvidas para o reconhecimento da espécie invasora *Acacia Longifolia*, permite atingir taxas de acerto e precisão mais elevadas, na função principal do classificador. O trabalho desenvolvido demonstrou também a utilidade das redes para tarefas de mapeamento da *Acacia Longifolia* em áreas de interesse, sem a necessidade da extração manual das características que é requerida pelos algoritmos existentes baseados em técnicas clássicas de aprendizagem automática. Portanto, a solução apresentada mostra maior potencial para fornecer melhores classificadores à medida que mais dados sejam disponibilizados.

## 6.1 Sugestões para Trabalho Futuro

A abordagem desenvolvida apresenta algumas limitações principalmente no método de segmentação através de uma janela deslizante com deslocamento de um pixel de cada vez, pois trata-se de um método demorado. A aplicação de redes neuronais completamente convolucionais (RCC) é uma forma mais eficiente para a

geração de imagens segmentadas automaticamente, ao invés da classificação através do método de redes neuronais convolucionais utilizado na abordagem proposta. Para tal, poder-se-ia recorrer a redes pré-treinadas e treiná-las para se ajustarem ao conjunto de dados. No entanto, é necessário realizar um conjunto de dados com a segmentação manual das imagens em grande escala para o treino de uma RCC.

Em relação ao conjunto de dados, poder-se-ia equilibrar as classes do conjunto de dados aplicando-se técnicas para equilibrar as classes através, por exemplo, da geração de amostras sintéticas e re-treinar a rede. Desta forma, contribui-se para a redução de sobre-ajuste da rede e o algoritmo não tende a aprender a classificar melhor uma dada classe por esta apresentar mais amostras. Para além disso, a rede desenvolvida foi treinada com amostras da espécie invasora *Acacia Longifolia* em época de floração o que limita o reconhecimento desta espécie fora desta época, pelo que seria importante capturar amostras da espécie sem flores e re-treinar a RNC.

Outros trabalhos futuros envolvem permitir que o veículo aéreo se aproxime das espécies detetadas, para uma inspeção mais detalhada e para o reconhecimento de outras espécies da família da Acácia e de outras plantas invasoras. Esta inspeção detalhada permitiria distinguir as várias espécies através da estrutura das folhas ou outras características para o reconhecimento das espécies invasoras. Em relação ao desempenho da rede, poder-se-ia recorrer ao estudo da variação dos hiper-parâmetros ou analisar outros algoritmos para o cálculo da medida de confiança. Poder-se-ia, ainda, aplicar o algoritmo desenvolvido a outros conjuntos de dados para se verificar o comportamento da curva da evolução da taxa de acerto melhorada em função da percentagem de amostras que foram revistas pelo especialista. Para além disso, poder-se-ia desenvolver um classificador automático composto por várias redes neuronais convolucionais com configurações de arquitetura e treino diferentes cujo resultado da previsão final seria a média das previsões obtidas por cada uma das redes neuronais convolucionais. Esta técnica permite aumentar a taxa de acerto, pois combina várias previsões diminuindo a variância dos modelos de redes neuronais convolucionais.

# Bibliografia

[Ammour et al., 2017] Ammour, N., Alhichri, H., Bazi, Y., Benjdira, B., Alajlan, N., and Zuair, M. (2017). Deep learning approach for car detection in uav imagery. *Remote Sensing*, 9(4):312.

[Audebert et al., 2017] Audebert, N., Le Saux, B., and Lefèvre, S. (2017). Segment-before-detect: Vehicle detection and classification through semantic segmentation of aerial images. *Remote Sensing*, 9(4):368.

[Badrinarayanan et al., 2017] Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495.

[Bergado et al., 2016] Bergado, J. R., Persello, C., and Gevaert, C. (2016). A deep learning approach to the classification of sub-decimetre resolution aerial images. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 1516–1519.

[Branco, 2017] Branco, E. P. (2017). *Cyberthreat discovery in open source intelligence using deep learning techniques*. PhD thesis, Ryersib University.

[Castelluccio et al., 2015] Castelluccio, M., Poggi, G., Sansone, C., and Verdoliva, L. (2015). Land use classification in remote sensing images by convolutional neural networks. *arXiv preprint arXiv:1508.00092*.

[Chollet, 2019] Chollet, F. (2015 (acedido Janeiro 1, 2019)). Keras. https://keras.io/.

[Cruzan et al., 2016] Cruzan, M. B., Weinstein, B. G., Grasty, M. R., Kohrn, B. F., Hendrickson, E. C., Arredondo, T. M., and Thompson, P. G. (2016). Small unmanned aerial vehicles (micro-uavs, drones) in plant ecology. *Applications in plant sciences*, 4(9).

[CS, 2019] CS, S. (2015 (acedido Julho 27, 2019)). Convolutional neural networks. http://cs231n.github.io/convolutional-networks/.

[de Sá et al., 2018] de Sá, N. C., Castro, P., Carvalho, S., Marchante, E., López-Núñez, F. A., and Marchante, H. (2018). Mapping the flowering of an invasive plant using unmanned aerial vehicles: is there potential for biocontrol monitoring? *Frontiers in plant science*, 9:293.

[Deng et al., 2018] Deng, Z., Sun, H., Zhou, S., Zhao, J., Lei, L., and Zou, H. (2018). Multi-scale object detection in remote sensing imagery with convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145:3–22.

[Dozat, 2016] Dozat, T. (2016). Incorporating nesterov momentum into adam.(2016). *Dostupné z: http://cs229. stanford. edu/proj2015/054_report. pdf*.

[Fan et al., 2018] Fan, Z., Lu, J., Gong, M., Xie, H., and Goodman, E. D. (2018). Automatic tobacco plant detection in uav images via deep neural networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(3):876–887.

[footage, 2019] footage, D. (2019 (acedido Julho 25, 2019)). Dji phantom 3 advanced. https://www.uavsystemsinternational.com/product/dji-phantom-3-advanced/.

[Gil et al., 2013] Gil, A., Lobo, A., Abadi, M., Silva, L., and Calado, H. (2013). Mapping invasive woody plants in azores protected areas by using very high-resolution multispectral imagery. *European Journal of Remote Sensing*, 46(1):289–304.

[Goodfellow et al., 2016] Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. http://www.deeplearningbook.org.

[Google, 2019] Google (2015 (acedido Janeiro 1, 2019)). Tensorflow. https://www.tensorflow.org/.

[Google, 2018] Google (Dezembro 1, 2018). Colaboratory. https://colab.research.google.com/notebooks/welcome.ipynb.

[Gray et al., 2019] Gray, P. C., Fleishman, A. B., Klein, D. J., McKown, M. W., Bézy, V. S., Lohmann, K. J., and Johnston, D. W. (2019). A convolutional neural network for detecting sea turtles in drone imagery. *Methods in Ecology and Evolution*, 10(3):345–355.

[Ham et al., 2018] Ham, S., Oh, Y., Choi, K., and Lee, I. (2018). Semantic segmentation and unregistered building detection from uav images using a deconvolutional network. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42:2.

[He et al., 2017] He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969.

[Hu and Ramanan, 2017] Hu, P. and Ramanan, D. (2017). Finding tiny faces. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 951–959.

[Huang et al., 2018] Huang, H., Deng, J., Lan, Y., Yang, A., Deng, X., and Zhang, L. (2018). A fully convolutional network for weed mapping of unmanned aerial vehicle (uav) imagery. *PloS one*, 13(4):e0196302.

[Intel, 2019] Intel (2000 (Janeiro 1, 2019)). Opencv. https://opencv.org/.

[Invasoras.pt, 2019] Invasoras.pt (2013 (Julho 25, 2019)). Acacia longifolia. http://invasoras.pt/gallery/acacia-longifolia/.

[Ioffe and Szegedy, 2015] Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.

[Kamilaris and Prenafeta-Boldú, 2018] Kamilaris, A. and Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and electronics in agriculture*, 147:70–90.

[Kaneko and Nohara, 2014] Kaneko, K. and Nohara, S. (2014). Review of effective vegetation mapping using the uav (unmanned aerial vehicle) method. *Journal of Geographic Information System*, 6(06):733.

[Kingma and Ba, 2014] Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

[LABS, 2019] LABS, S. (2013 (acedido Janeiro 1, 2019)). Understanding deep learning: Dnn, rnn, lstm, cnn and r-cnn. https://mc.ai/understanding-deep-learning-dnn-rnn-lstm-cnn-and-r-cnn/.

[LeCun et al., 1999] LeCun, Y., Haffner, P., Bottou, L., and Bengio, Y. (1999). Object recognition with gradient-based learning. In *Shape, contour and grouping in computer vision*, pages 319–345. Springer.

[Liu et al., 2016] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A. C. (2016). Ssd: Single shot multibox detector. In *European conference on computer vision*, pages 21–37. Springer.

[Liu et al., 2018] Liu, Y., Sun, P., Highsmith, M. R., Wergeles, N. M., Sartwell, J., Raedeke, A., Mitchell, M., Hagy, H., Gilbert, A. D., Lubinski, B., et al. (2018). Performance comparison of deep learning techniques for recognizing birds in aerial images. In *2018 IEEE Third International Conference on Data Science in Cyberspace (DSC)*, pages 317–324.

[Long et al., 2015] Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440.

[Lopatin et al., 2019] Lopatin, J., Dolos, K., Kattenborn, T., and Fassnacht, F. E. (2019). How canopy shadow affects invasive plant species classification in high spatial resolution remote sensing. *Remote Sensing in Ecology and Conservation*, pages 1–16.

[Maggiori et al., 2017] Maggiori, E., Tarabalka, Y., Charpiat, G., and Alliez, P. (2017). High-resolution aerial image labeling with convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 55(12):7092–7103.

[Mallick, 2019] Mallick, S. (2017 (acedido Agosto 20, 2019)). Bias-variance tradeoff in machine learning. https://www.learnopencv.com/bias-variance-tradeoff-in-machine-learning/?fbclid=IwAR0mTM1xkWD_YPlWOIaHLdEpbi7ER1exU9LCdpZukuflJGa9qg_FwlNKskM.

[Marchante et al., 2014] Marchante, H., Morais, M., Freitas, H., and Marchante, E. (2014). *Guia prático para a identificação de Plantas Invasoras em Portugal*. Imprensa da Universidade de Coimbra/Coimbra University Press.

[Martins, 2012] Martins, F. D. (2012). *Utilização de técnicas de deteção remota na identificação de Acacia sp. na Região Centro Sul de Portugal Continental*. PhD thesis, IPCB. ESA.

[Mendes and Dal Poz, 2011] Mendes, T. S. G. and Dal Poz, A. P. (2011). Classificação de imagens aéreas de alta-resolução utilizando redes neurais artificiais e dados de varredura a laser. *SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO*, 15:7792–7799.

[Murphy, 2016] Murphy, J. (2016). An overview of convolutional neural network architectures for deep learning. https://www.microway.com/download/whitepaper/An_Overview_of_Convolutional_Neural_Network_Architectures_for_Deep_Learning_fall2016.pdf.

[Najibi et al., 2017] Najibi, M., Samangouei, P., Chellappa, R., and Davis, L. S. (2017). Ssh: Single stage headless face detector. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4875–4884.

[Noh et al., 2015] Noh, H., Hong, S., and Han, B. (2015). Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE international conference on computer vision*, pages 1520–1528.

[Onishi and Ise, 2018] Onishi, M. and Ise, T. (2018). Automatic classification of trees using a uav onboard camera and deep learning. *arXiv preprint arXiv:1804.10390*.

[Pande-Chhetri et al., 2017] Pande-Chhetri, R., Abd-Elrahman, A., Liu, T., Morton, J., and Wilhelm, V. L. (2017). Object-based classification of wetland vegetation using very high-resolution unmanned air system imagery. *European Journal of Remote Sensing*, 50(1):564–576.

[Patterson and Gibson, 2017] Patterson, J. and Gibson, A. (2017). *Deep learning: A practitioner's approach*. "O'Reilly Media, Inc.".

[Paz-Kagan et al., 2019] Paz-Kagan, T., Silver, M., Panov, N., and Karnieli, A. (2019). Multispectral approach for identifying invasive plant species based on flowering phenology characteristics. *Remote Sensing*, 11(8):953.

[Qian, 1999] Qian, N. (1999). On the momentum term in gradient descent learning algorithms. *Neural networks*, 12(1):145–151.

[Radovic et al., 2017] Radovic, M., Adarkwa, O., and Wang, Q. (2017). Object recognition in aerial images using convolutional neural networks. *Journal of Imaging*, 3(2):21.

[Redmon et al., 2016] Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788.

[Ren et al., 2015] Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99.

[Ronneberger et al., 2015] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.

[Sa et al., 2017] Sa, I., Chen, Z., Popović, M., Khanna, R., Liebisch, F., Nieto, J., and Siegwart, R. (2017). weednet: Dense semantic weed classification using multispectral images and mav for smart farming. *IEEE Robotics and Automation Letters*, 3(1):588–595.

[Sa et al., 2018] Sa, I., Popović, M., Khanna, R., Chen, Z., Lottes, P., Liebisch, F., Nieto, J., Stachniss, C., Walter, A., and Siegwart, R. (2018). Weedmap: a large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming. *Remote Sensing*, 10(9):1423.

[Safonova et al., 2019] Safonova, A., Tabik, S., Alcaraz-Segura, D., Rubtsov, A., Maglinets, Y., and Herrera, F. (2019). Detection of fir trees (abies sibirica) damaged by the bark beetle in unmanned aerial vehicle images with deep learning. *Remote Sensing*, 11(6):643.

[Sameen et al., 2018] Sameen, M. I., Pradhan, B., and Aziz, O. S. (2018). Classification of very high resolution aerial photos using spectral-spatial convolutional neural networks. *Journal of Sensors*, 2018.

[Ševo and Avramović, 2016] Ševo, I. and Avramović, A. (2016). Convolutional neural network based automatic object detection on aerial images. *IEEE geoscience and remote sensing letters*, 13(5):740–744.

[Simonyan and Zisserman, 2014] Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[Srivastava et al., 2014] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958.

[Sun et al., 2018] Sun, Y., Zhang, X., Xin, Q., and Huang, J. (2018). Developing a multi-filter convolutional neural network for semantic segmentation using high-resolution aerial imagery and lidar data. *ISPRS journal of photogrammetry and remote sensing*, 143:3–14.

[Tang et al., 2017] Tang, T., Zhou, S., Deng, Z., Lei, L., and Zou, H. (2017). Fast multidirectional vehicle detection on aerial images using region based convolutional neural networks. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 1844–1847.

[Tudorache et al., 2017] Tudorache, S., Popescu, D., and Ichim, L. (2017). Combining efficient textural features with cnn—based classifiers to segment regions of interest in aerial images. In *2017 5th International Symposium on Electrical and Electronics Engineering (ISEEE)*, pages 1–6.

[Van den Bergh et al., 2012] Van den Bergh, M., Boix, X., Roig, G., de Capitani, B., and Van Gool, L. (2012). Seeds: Superpixels extracted via energy-driven sampling. In *European conference on computer vision*, pages 13–26. Springer.

[Wang et al., 2019] Wang, S., Liu, L., Qu, L., Yu, C., Sun, Y., Gao, F., and Dong, J. (2019). Accurate ulva prolifera regions extraction of uav images with superpixel and cnns for ocean environment monitoring. *Neurocomputing*, 348:158–168.

[Wen and Geomatics, 2016] Wen, R. and Geomatics, H. (2016). Geospatial semantic pattern recognition in volunteered geographic data using the random forest algorithm. Master's thesis.

[Zaragoza and d'Alché Buc, 1998] Zaragoza, H. and d'Alché Buc, F. (1998). Confidence measures for neural network classifiers. In *Proceedings of the Seventh Int. Conf. Information Processing and Management of Uncertainty in Knowlegde Based Systems*.