



INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Segmentação de Imagens de Zonas Urbanas em Espaços Florestais

Diogo Alexandre Ferreira Dinis

Mestrado em Sistemas Integrados de Apoio à Decisão

Orientador:

Doutor Tomás Gomes da Silva Serpa Brandão, Professor Auxiliar

ISCTE - Instituto Universitário e Lisboa

Co-Orientador:

Doutor João Carlos Amaro Ferreira, Professor Auxiliar com Agregação

ISCTE - Instituto Universitário e Lisboa

Novembro, 2022



TECNOLOGIAS
E ARQUITETURA

Departamento de Ciências e Tecnologias da Informação

Segmentação de Imagens de Zonas Urbanas em Espaços Florestais

Diogo Alexandre Ferreira Dinis

Mestrado em Sistemas Integrados de Apoio à Decisão

Orientador:

Doutor Tomás Gomes da Silva Serpa Brandão, Professor Auxiliar

ISCTE - Instituto Universitário e Lisboa

Co-Orientador:

Doutor João Carlos Amaro Ferreira, Professor Auxiliar com Agregação

ISCTE - Instituto Universitário e Lisboa

Novembro, 2022

Direitos de cópia ou Copyright

©Copyright: Diogo Alexandre Ferreira Dinis

O Instituto Universitário de Lisboa (ISCTE-IUL) tem o direito, perpétuo e sem limites geográficos, de arquivar e publicitar este trabalho através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, de o divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

Agradecimentos

A entrega deste projeto de dissertação simboliza um momento muito importante na minha vida, significando a finalização de uma fase e o início de uma nova. Ao longo deste projeto várias dificuldades e desafios foram surgindo, no entanto todas elas foram ultrapassadas com o apoio das pessoas à minha volta.

Gostaria de agradecer em primeiro lugar aos meus orientadores, professor João Ferreira e professor Tomás Brandão, por todo o apoio e disponibilidade que foram tendo ao longo deste projeto, onde mesmo quando poderia estar menos motivado, foram puxando por mim para a conclusão desta dissertação.

Quero fazer um agradecimento especial à minha família, por todo o apoio que me foi dado durante os meus anos de estudo e um agradecimento especial ao meu grupo de amigos mais próximo, Carolina, Joana, João Rui, Paulo e Sara, por todos os momentos de descontração, que foram importantes nesta fase da minha vida académica e pessoal. Gostaria de fazer um agradecimento também aos meus amigos Pedro, Chora e Gonçalo, que fizeram parte da minha vida académica e pessoal, e que sem eles não teria conseguido chegar a esta fase.

Por último gostaria de agradecer a todas as pessoas envolvidas durante esta fase da minha vida, pela compreensão e incentivo que foram dado ao longo da realização desta dissertação, e que sem eles não seria possível a sua conclusão.

Resumo

Os incêndios florestais são das catástrofes naturais mais graves, não apenas pela frequência com que estes acontecem e dimensão que alcançam, como também pelos efeitos devastadores económicos e ambientais.

Nos últimos anos, Portugal não tem tido a capacidade de travar eficazmente estes fenómenos naturais. Torna-se por isso necessário concentrar esforços na prevenção e na deteção, de forma que se consiga uma ação mais rápida sobre os focos de incêndios florestais assim que estes emergem.

O INOV (Instituto de Engenharia de Sistemas e Computadores Inovação) desenvolveu um sistema de televigilância – CICLOPE – para monitorização de incêndios florestais através de torres de videovigilância que recolhem imagens. Entre as funcionalidades implementadas, existe um sistema responsável pela deteção automática de incêndios. Contudo, para esse sistema funcionar corretamente, as imagens recolhidas necessitam de estar devidamente segmentadas entre espaços florestais e zonas urbanas. Esta segmentação é realizada manualmente pelos profissionais do INOV, sendo um processo longo e suscetível a erros.

Com o intuito de automatizar o processo de segmentação, nesta dissertação foi elaborada uma base de dados recorrendo às imagens fornecidas pelo INOV. Após a análise dos dados obtidos foram estudadas várias abordagens de segmentação, sendo selecionadas três que melhor se adequavam ao problema. As abordagens selecionadas foram, *thresholding*, K-Means e a arquitetura DeepLabv3, sendo as últimas duas referentes à vertente de aprendizagem automática. Após uma comparação dos resultados dos três modelos estudados, a arquitetura DeepLabv3 exibiu o melhor resultado, apresentando os valores médios de Intersection over Union e Dice Coefficient 0,51 e 0,78, respetivamente.

Palavras-Chave: Segmentação de imagem, Incêndios Florestais, Visão por Computador, Redes Neurais Convolucionais

Abstract

Forest fires are among the most serious natural disasters, not only because of their frequency and scale, but also because of their devastating economic and environmental effects.

In the last years, Portugal has not had the capacity to effectively stop this natural phenomenon. It is therefore necessary to concentrate efforts on prevention and detection, so that faster action can be taken on forest fire outbreaks as soon as they emerge.

INOV (Instituto de Engenharia de Sistemas e Computadores Inovação) has developed a surveillance system - CICLOPE - for monitoring forest fires through video surveillance towers that collect images. Among the implemented functionalities, there is a system responsible for automatic fire detection. However, for this system to work properly, the collected images need to be properly segmented between forest spaces and urban areas to minimize the number of false positives. This segmentation is performed manually by INOV professionals and is a long process and susceptible to errors.

With the aim of automating the segmentation process, in this dissertation a database was developed using the images provided by INOV. After the analysis of the obtained data, several approaches of segmentation were studied, being selected three that best suited the problem. The approaches selected were thresholding, K-Means and DeepLabv3 architecture, the last two referring to machine learning. After a comparison of the results of the three models studied, the DeepLabv3 architecture exhibited the best result, presenting the average values of Intersection over Union and Dice Coefficient 0,51 and 0,78, respectively.

Keywords: Image Segmentation, Wildfires, Computer Vision, Convolutional Neural Network

Índice

Agradecimentos	i
Resumo	iii
Abstract	v
Índice	vii
Índice de Tabelas	ix
Índice de Figuras	xi
Capítulo 1 – Introdução	1
1.1. Enquadramento do tema	1
1.2. Motivação e relevância do tema	3
1.3. Questões e objetivos de investigação	3
1.3.1. Questão de investigação	3
1.3.2. Objetivos	3
1.4. Abordagem metodológica	4
1.5. Estrutura e organização da dissertação	5
Capítulo 2 – Revisão da Literatura	7
2.1. Algoritmos de segmentação de imagem	7
2.2. Técnicas clássicas de segmentação de imagem	11
2.2.1. Segmentação por contornos	11
2.2.2. Segmentação por região	12
2.2.3. Segmentação por <i>thresholding</i>	13
2.3. Técnicas de aprendizagem automática de segmentação de imagem	15
2.3.1. <i>K-Means</i>	15
2.3.2. Redes Neurais Convolucionais	16
2.4. Processos de tratamento de imagem	18
Capítulo 3 – Análise e Exploração dos Dados	21
3.1. Conjunto de Dados	21

3.2. Tratamento do conjunto de dados	23
Capítulo 4 – Desenvolvimento	25
4.1. Thresholding	26
4.2. <i>K-Means</i>	26
4.3. DeepLabv3	27
4.3.1. Tratamento dos dados	27
4.3.2. Arquitetura do modelo DeepLaV3	28
4.3.3. Treino e Avaliação	28
Capítulo 5 – Análise e Avaliação dos Resultados	31
5.1. Resultados dos modelos	31
5.1.1. Resultados do método por <i>thresholding</i>	31
5.1.2. Resultados do modelo <i>K-Means</i>	33
5.1.3. Resultados do modelo DeepLabV3	35
5.2. Avaliação dos resultados	36
Capítulo 6 – Conclusão	39
6.1. Limitações do trabalho	41
6.2. Trabalho futuro	41
Referências	43

Índice de Tabelas

Tabela 1: Performance de diversos métodos sobre a base de dados Cityscapes [17].	17
Tabela 2: Tabela de dados.....	21
Tabela 3: Comparação de resultados.....	35
Tabela 4: Resultados dos métodos de avaliação	38

Índice de Figuras

Figura 1: Metodologia DSRM adaptada	iv
Figura 2: Técnicas de segmentação de imagem [5]	8
Figura 3: Técnicas de segmentação de imagem e as suas vantagens e desvantagens [6]	9
Figura 4: Distribuição dos pontos de dados [12]	10
Figura 5: K-Means [13] vs. C-Means [14].....	10
Figura 6: Técnicas de segmentação por contornos [5].....	11
Figura 7: Segmentação por contornos, Canny Edge Detection [7].....	12
Figura 8: Implementação da técnica de segmentação por região [18].....	13
Figura 9: Imagens de zona de floresta e zona citadina [8].....	13
Figura 10: Segmentação de zona de floresta e zona citadinas [8]	14
Figura 11: Exemplo de imagens no espaço de cor HSV [12].	14
Figura 12: Exemplo de imagens segmentadas [12]	14
Figura 13: Aplicação do algoritmo K-Means a imagens de leucemia [16]	15
Figura 14: Processo da arquitetura DeepLabv3 [43]	18
Figura 15: Segmentação de imagem através do método DeepLabV3 [24]	18
Figura 16: Exemplo antes e depois da utilização do algoritmo de remoção de nevoeiro [45]	19
Figura 17: Distribuição das imagens por zonas de florestação.....	22
Figura 18: Imagens e respetivas máscaras fornecidas pelo INOV	23
Figura 19: Subimagens realizadas às imagens e máscaras fornecidas.....	23
Figura 20: Tratamento da imagem quanto ao ruído.....	24
Figura 21: Esquema de remoção de clusters	27
Figura 22: Cálculo do método de avaliação Intersection over Union [37]	29
Figura 23: Cálculo do método de avaliação Dice Coefficient [17]	29
Figura 24: Máscara por thresholding em diversas alturas do dia.....	32
Figura 25: Máscara por Thresholding, a) imagem original, b) máscara criada, c) máscara manual	32
Figura 26: Resultados positivos K-Means, a) imagem original, b) máscara criada, c) máscara manual	34
Figura 27: Resultados negativos K-Means, a) imagem original, b) máscara criada, c) máscara manual	34
Figura 28: Resultados DeepLabV3	35
Figura 29: Avaliação do modelo DeepLabV3 recorrendo às subimagens realizadas.....	37
Figura 30: Avaliação do modelo DeepLabV3 recorrendo às imagens com tamanho original.....	38

Introdução

1.1. Enquadramento do tema

Todos os anos em Portugal extensas áreas florestais são destruídas, apresentando uma das mais elevadas taxas de ocorrências de fogos florestais a nível mundial, sendo 98% destas ocorrências causadas em 2022 (dados de 1 de Janeiro a 22 de Outubro) por intervenção humana, seja ela por negligência ou por mão criminosa, segundo o ICNF (Instituto da Conservação da Natureza e Florestas) [1].

Apesar de todos os avanços científicos, os fogos florestais continuam a surgir todos os anos de forma cada vez mais acentuada. Para um combate mais eficaz torna-se imperativo uma deteção imediata antes que estes focos de incêndio se alastrem, de forma a facilitar a sua extinção.

Vários métodos foram desenvolvidos para a deteção de fogos florestais emergentes, como a instalação de sistemas de videovigilância em espaços florestais, sendo as imagens recebidas por agentes da proteção civil que fazem a sua monitorização e análise, procurando detetar possíveis focos de incêndio. No entanto, devido à extensa área florestal portuguesa e aos poucos recursos humanos disponibilizados para a sua monitorização, torna-se imperativo um sistema de deteção automática que auxilie os agentes destacados para a localização de incêndios.

Com esse intuito, o INOV (Instituto de Engenharia de Sistemas e Computadores Inovação) desenvolveu o sistema de televigilância CICLOPE. O CICLOPE foi concebido para ser um sistema de monitorização de incêndios florestais com deteção automática de incêndios emergentes e alerta de primeira resposta instantânea, permitindo identificar colunas de fumo, pontos quentes ou chamas ao longo de 24 horas, mesmo sob as condições mais adversas. [2].

Uma das componentes do CICLOPE é o vídeo e a imagem recolhida a partir de torres de vigilância. Existem diversos tipos de torres de vigilância dispersas pelas florestas portuguesas. No entanto, nem todas têm a capacidade de obtenção de dados. No caso do CICLOPE, pelas características implementadas, este permite monitorizar remotamente grandes áreas de terreno, a qualquer altura do dia. Através da utilização simultânea ou individual de câmaras de vídeo da gama do visível e dos infravermelhos, assim como LIDAR (Laser Imaging, Detection, and Ranging), são efetuadas observações diurnas e noturnas em quase todas as condições meteorológicas [2].

Todo o equipamento necessário a esta análise e recolha de imagens é instalado sobre as torres denominadas TVAD (Torres de Vigilância e Aquisição de Dados), podendo ser controladas a partir de diferentes locais como um quartel de Bombeiros, instalação da Proteção Civil, posto de Polícia, entre outros.

Estas torres de vigilância surgem com o intuito de suprimir lacunas existentes quanto à vigilância exercida por parte dos agentes responsáveis, uma vez que estes muitas das vezes não conseguem alcançar locais de difícil acesso, e não abrangem grande parte do espaço florestal. As torres referidas são então vistas como uma alternativa ao combate aos incêndios, obtendo informação acerca dos mesmos e fornecendo a localização exata destes.

No entanto, não basta a recolha de imagens para um eficaz combate aos incêndios, é necessária uma deteção automática destes focos de incêndio, e para isso o INOV desenvolveu um algoritmo que tem como base a deteção de fumo em espaços florestais. Esta deteção tem o auxílio de máscaras previamente desenhadas manualmente que se sobrepõem às imagens recolhidas. Na construção das máscaras são tidas em conta as zonas urbanas e os espaços florestais existentes, sendo coberta toda a zona urbana uma vez que o objetivo é a deteção de incêndios em espaços florestais, tentando evitar desta forma o surgimento de alertas falsos positivos.

Contudo, o aparecimento de alertas falsos positivos é recorrente, levando a uma sobrecarga dos agentes da proteção civil ao validar cada um destes alarmes. A existência de vários falsos positivos leva ainda à redução de credibilidade do sistema, sendo posto em causa o seu funcionamento por parte dos recursos humanos destacados para exercerem a sua monitorização. Torna-se assim necessária uma redução considerável destas ocorrências.

Após uma análise destes alertas verificou-se que grande parte teriam origem nas máscaras criadas para a utilização do algoritmo de deteção automática de incêndios. As máscaras por vezes estariam mal desenhadas, não cobrindo a totalidade da zona urbana, levando ao algoritmo a identificar zonas urbanas como possíveis focos de incêndio. Sendo então uma área de interesse a automatização destas máscaras, de forma a ser possível distinguir entre zonas urbanas e espaços florestais, auxiliando a construção destas, reduzindo a necessidade de intervenção humana.

A recolha de imagens retratada será material de estudo desta dissertação, com o objetivo de otimizar o processo desenvolvido pelo INOV através da segmentação automática das imagens recolhidas, tentando desta forma uma redução do número de falsos positivos através da automatização do processo de segmentação manual já existente.

1.2. Motivação e relevância do tema

A escolha deste tema assentou em duas motivações sendo elas, o meu interesse pessoal pela área da automatização e a relevância que este tema tem na nossa sociedade.

As zonas afetadas em Portugal pelos incêndios florestais são diversas, e muitas delas são de difícil acesso, dificultando o combate antecipado aos focos de incêndio. A esta temática une-se evolução tecnológica, o que torna claro a extrema importância que a automatização de processos tem neste combate em alturas mais críticas do ano. Atualmente o INOV já assegura uma identificação dos focos de incêndios através de diversas torres de vigilância. A automatização de alguns processos existentes já referidos, de forma a melhorá-los, pode tornar possível um progresso no combate aos incêndios em Portugal.

1.3. Questões e objetivos de investigação

1.3.1. Questão de investigação

A questão de investigação estabelecida como ponto de partida foi: *De que forma é possível automatizar a segmentação de espaços florestais através de técnicas visão por computador, e como pode esta contribuir para a melhoria do sistema de combate aos incêndios emergentes?*

1.3.2. Objetivos

Esta dissertação tem como principal objetivo a criação de um protótipo que permita a marcação automática das zonas urbanas em espaços florestais. O caso de estudo em causa é focado no combate aos incêndios emergentes em espaços florestais, incêndios estes, que se não forem combatidos o quanto antes, poderão levar a uma série de catástrofes, tanto no ecossistema como na economia de Portugal.

Pretende-se explorar diversas técnicas de visão por computador, assim como técnicas de aprendizagem automática com vista à identificação de zonas urbanas em espaços florestais, automatizando assim a criação das máscaras que são criadas atualmente pelo INOV num processo manual. Esta investigação tem também como finalidade a demonstração do benefício que as técnicas representadas nesta dissertação podem trazer para uma melhor eficácia nos combates aos incêndios.

1.4. Abordagem metodológica

Esta dissertação é realizada com o apoio do INOV, sendo este o fornecedor dos dados necessários à sua realização. Foi utilizado um conjunto de imagens de diversos espaços florestais, e respetivas máscaras elaboradas pelo INOV, máscaras estas que cobrem as zonas urbanas a eliminar. De forma a perceber e avaliar o protótipo elaborado, comparam-se os resultados obtidos através do processo de automatização com as máscaras realizadas manualmente pelo INOV.

Após uma análise do problema existente, decidiu-se que a metodologia DSRM [3] seria adequada à resolução do mesmo. No entanto, optou-se por adaptar esta adicionando os campos “Compreensão e preparação dos dados”, “Sugestão” e “Conclusão”, como exemplificado na Figura 1. O primeiro campo adicionado tem em vista a análise das imagens recebidas, assim como a sua preparação de forma a ser possível ter uma ideia mais rigorosa do material em estudo e a uma escolha mais acertada da abordagem. O segundo campo tem como objetivo a análise e seleção das abordagens, com o intuito de entender quais as mais adequadas perante o entendimento do problema. Por último, o terceiro campo adicionado, denominado “Conclusão”, pretende sintetizar as escolhas realizadas ao longo da metodologia, assim como os resultados finais.

Esta dissertação irá seguir as mesmas etapas que a metodologia referida, sendo assim este trabalho dividido em 6 fases:

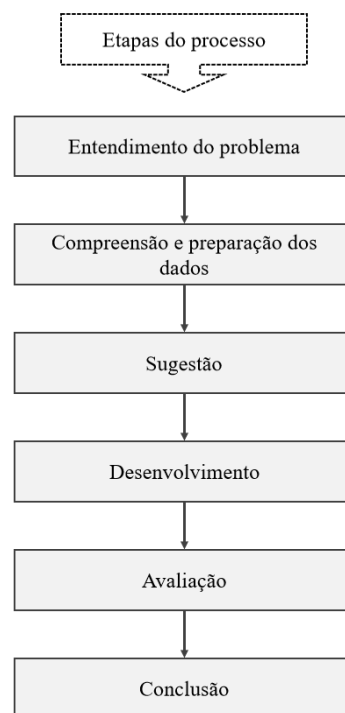


Figura 1: Metodologia DSRM adaptada

Entendimento do problema – Numa primeira fase será necessário fazer uma análise ao problema em questão, recolhendo todos os detalhes sobre este, o que é realizado atualmente e quais os objetivos a serem alcançados.

Compreensão e preparação dos dados – Numa segunda fase será realizada uma análise dos dados fornecidos pelo INOV, com o objetivo de se perceber a sua estrutura. Em seguida, proceder-se-á ao tratamento dos dados com vista à extração de algum ruído existente nas imagens como nevoeiro, divisão de imagens em pequenos segmentos, entre outros.

Sugestão – Com os dados tratados seguir-se-á a análise das abordagens selecionadas para o desenvolvimento do protótipo, onde serão descritas individualmente as suas características, como vantagens e desvantagens, e expectativas finais.

Desenvolvimento – Com as abordagens retratadas será então procedida a realização do processo de segmentação, aplicando diversas técnicas de visão por computador. Este processo de segmentação contará com técnicas mais clássicas de segmentação de imagem, assim como técnicas de aprendizagem automática, como a aplicação de um modelo de redes neuronais convolucionais. Este modelo será treinado com imagens de espaços florestais e respetivas máscaras que correspondem às zonas urbanas a segmentar,

Avaliação – Após a implementação dos modelos de segmentação abordados será realizada uma avaliação dos resultados dos mesmos. Esta avaliação terá em vista a análises dos resultados através de diversos métodos de avaliação e análise das máscaras obtidas, comparando também com os objetivos propostos.

Conclusão – Por último, e uma vez realizada a avaliação referida no ponto anterior, serão tiradas as devidas conclusões sobre as abordagens aplicadas.

1.5. Estrutura e organização da dissertação

A presente dissertação encontra-se organizada em 6 capítulos, sendo refletido em cada um deles as diferentes fases da investigação.

No **Capítulo 1** pretende-se dar a conhecer o tema em questão, motivações, objetivos que se pretendem obter e as abordagens que serão executadas ao longo do projeto.

O **Capítulo 2** apresenta uma revisão da literatura e tem como objetivo expor as metodologias utilizadas, e a pesquisa que suporta as escolhas realizadas. Realizando assim uma análise de trabalhos realizados anteriormente que estão relacionados com o tema da dissertação.

No **Capítulo 3** vai-se de encontro ao segundo passo da metodologia apresentada (Compreensão e preparação dos dados), sendo realizada a análise dos dados fornecidos pelo INOV, e o seu respetivo tratamento, de forma a otimizar a aplicação de técnicas de visão por computador, assim como das redes neuronais convolucionais.

Em seguida, no **Capítulo 4** serão retratadas as técnicas de visão por computador e redes neuronais convolucionais referidas no Capítulo 2 que mais se adequam à resolução do problema, e por conseguinte postas estas em prática, aplicando a metodologia referida anteriormente.

O **Capítulo 5** tem como finalidade apresentar os resultados obtidos e comparar estes com os resultados já existentes, exibindo as vantagens e desvantagens de cada um deles. Podendo verificar-se qual o método adequado tendo em conta as necessidades existentes.

Por último, o **Capítulo 6** apresenta os resultados alcançados com a dissertação e o que poderá ser aprimorado. Será ainda realizada a verificação dos objetivos propostos face aos alcançados, como ainda as adversidades encontradas ao longo da dissertação.

Revisão da Literatura

“Computer vision and machine learning have really started to take off, but for most people, the whole idea of what is a computer seeing when it's looking at an image is relatively obscure.”

[4]

(Mike Krieger)

Neste capítulo trata-se o tema da segmentação de imagem, apresentando os diversos tipos de segmentação de imagem, as suas características e importância para o tópico em questão. Trata-se ainda a forma como esta segmentação pode ser implementada, recorrendo a técnicas clássicas de visão por computador assim como técnicas de aprendizagem automática. Tendo em conta o objetivo suportado por esta dissertação, é abordado o método atual de deteção de incêndios desenvolvido pelo INOV e como este pode ser melhorado recorrendo às técnicas referidas anteriormente.

2.1. Algoritmos de segmentação de imagem

Ao longo desta dissertação serão exploradas imagens retiradas das torres de vigilância pertencentes ao INOV com o intuito de criar máscaras que possibilitem a distinção entre zonas urbanas e espaços florestais. A segmentação de imagem tem como objetivo dividir uma imagem em várias regiões ou objetos segundo um critério estabelecido, existindo diversas técnicas para a obtenção deste resultado divididas em dois grandes grupos (Figura 2): técnicas clássicas de segmentação, que recorrem a processos de tratamento sobre a imagem recebida; e técnicas baseadas em aprendizagem automática. Devido ao referido, existiu a necessidade de investigar as técnicas de segmentação de imagens existentes, as suas vantagens e desvantagens e projetos que já teriam sido realizados com estas, com vista a uma escolha de quais seriam as técnicas mais promissoras para a realização deste protótipo de automatização.

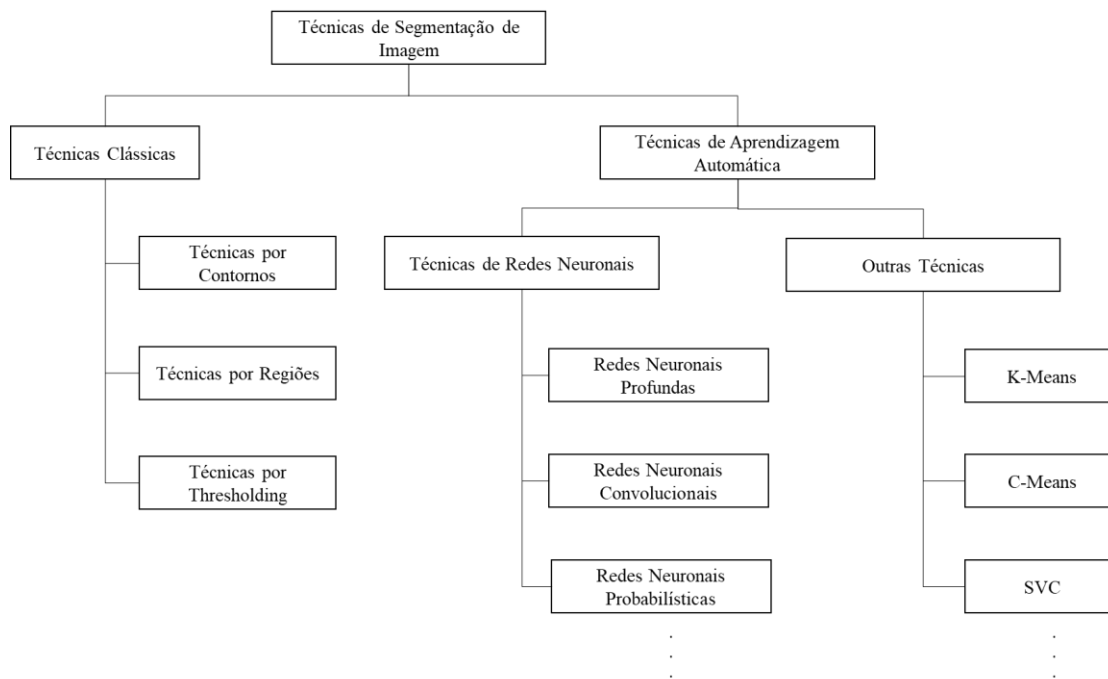


Figura 2: Técnicas de segmentação de imagem [5]

As técnicas clássicas de segmentação de imagem podem ser divididas segundo três grupos: por contornos, por regiões, e por *thresholding* [6], como ilustrado na Figura 3.

A primeira técnica baseia-se na segmentação de uma imagem através do reconhecimento de linhas, contornos ou pontos isolados, identificando mudanças abruptas nestes atributos, sendo os limites das regiões inferidas posteriormente [7]. Esta técnica torna-se bastante eficaz em imagens em que os contornos existentes são bastante destacados, permitindo realizar facilmente uma segmentação da imagem desejada. No entanto, em imagens onde existem demasiados contornos, ou em que estes não sejam perceptíveis, esta técnica acaba por não ser eficaz.

Quanto à segunda técnica, por regiões, esta recorre à homogeneidade da informação existente na imagem, tal como textura, intensidade e cor. Neste tipo de segmentação é necessário um pré-estabelecimento de regras, que deverão ser obedecidas, de forma que os pixels sejam classificados consoante as características de uma determinada região [5]. Esta técnica é mais adequada em imagens em que os contornos possam ser imperceptíveis, no entanto é bastante suscetível a variações dos resultados consoante os elementos dados para a realização da segmentação.

Na terceira técnica, por *thresholding*, é calculado um valor ótimo de *threshold* que irá separar duas classes, sendo posteriormente cada pixel atribuído a uma classe definida [8]. Pela sua simplicidade este método encontra-se entre os mais utilizados, no entanto não é um método

eficaz quanto à segmentação de imagens com um nível de informação elevado, sendo sobretudo usado na separação de objetos do plano de fundo.

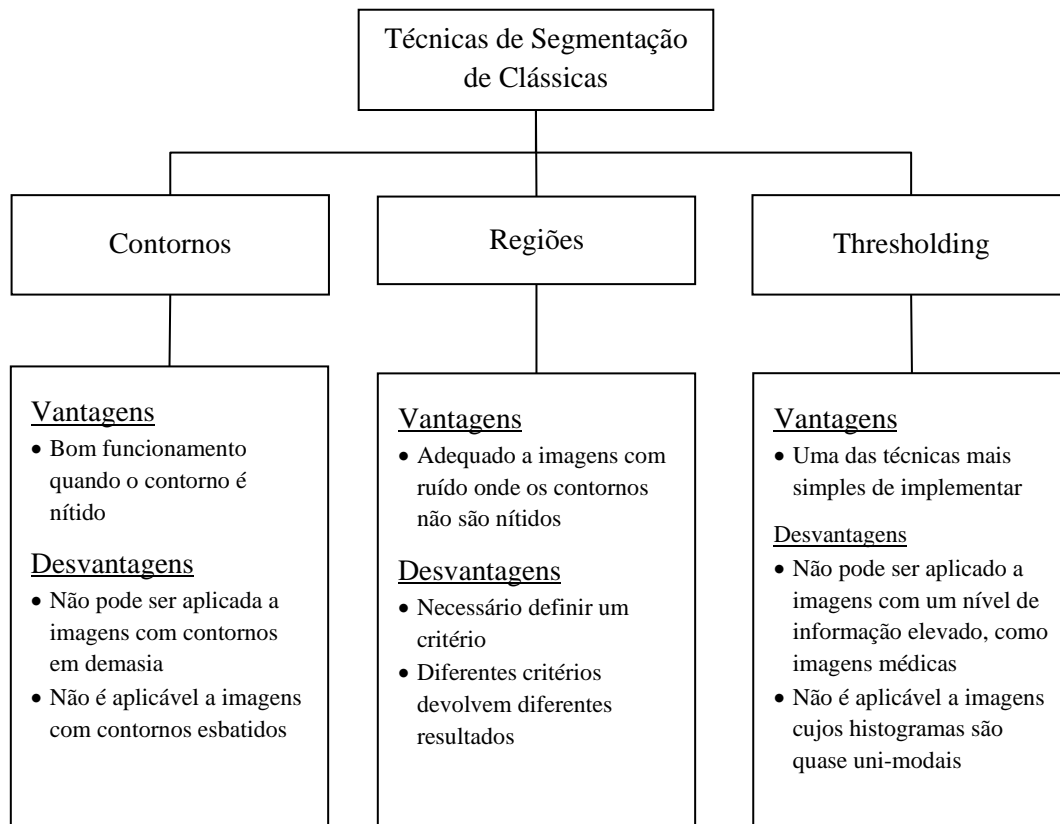


Figura 3: Técnicas de segmentação de imagem e as suas vantagens e desvantagens [6]

Para além das técnicas clássicas de segmentação de imagem existem outras como técnicas recorrendo a aprendizagem automática, estas podem ser construídas com base em redes neuronais ou não, como ilustrado na Figura 2.

As redes neuronais convolucionais (RNCs) serão o principal foco desta dissertação dentro da categoria de “Técnicas de Redes Neuronais”. Estas conseguiram elevar o desempenho dos sistemas de visão por computador numa vasta gama de problemas com alto nível de detalhe, incluindo a classificação de imagens e a deteção de objetos. Sendo as RNCs, quando treinadas, capazes de originar resultados impressionantemente melhores comparados aos sistemas que se baseiam em características artesanais (técnicas clássicas), evidenciados anteriormente. Essencial para este sucesso é a invariância integrada das RNCs às transformações de imagem locais, o que lhes permite aprender representações de dados cada vez mais abstratas [9]. Esta invariância é claramente desejável para tarefas de classificação, mas pode dificultar tarefas de previsão mais complexas, tais como a segmentação semântica, onde a abstração de informação espacial é indesejável.

Quanto às outras técnicas de aprendizagem automática é possível destacar as técnicas por *clusters*, K-Means e C-Means. Como observado na Figura 4, na primeira técnica, categorizada como “Hard Clustering”, os pontos de dados são divididos em dois *clusters* e cada ponto pertencente a um dos dois aglomerados [11]. Já a segunda técnica categoriza-se como “Soft Clustering”, onde em vez de se categorizar cada ponto de dados em *clusters* separados, é atribuída uma probabilidade de esse ponto pertencer a um determinado *cluster* [11].

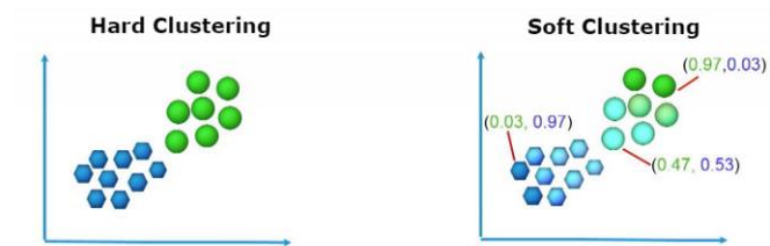


Figura 4: Distribuição dos pontos de dados [12]

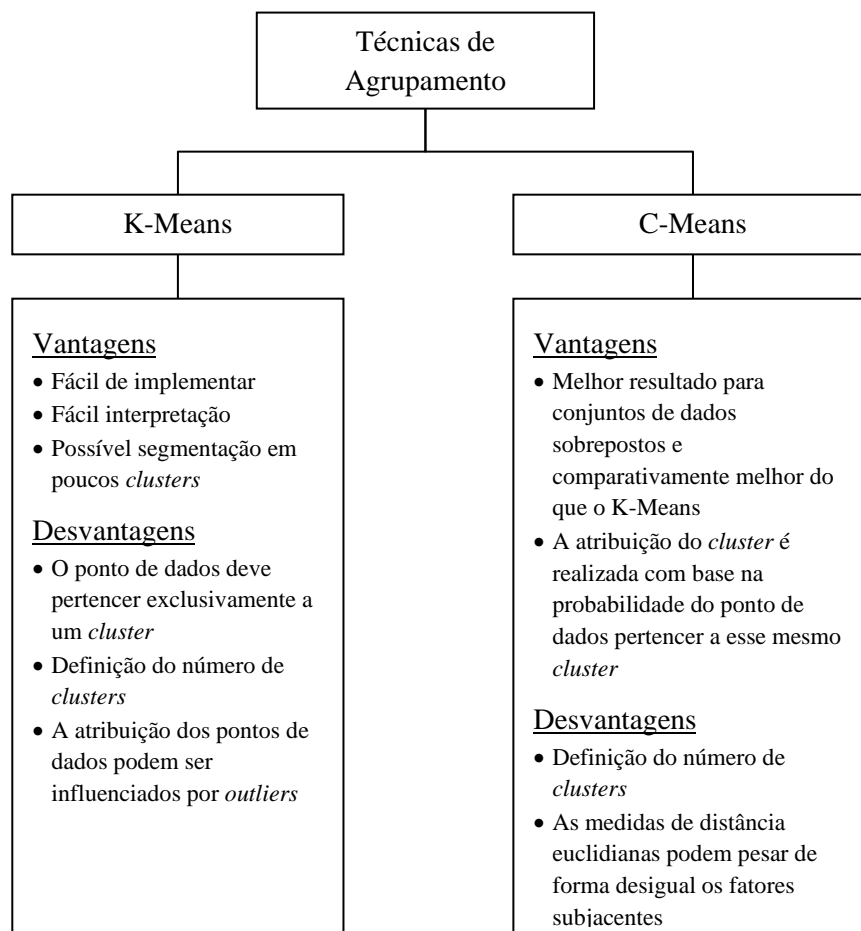


Figura 5: K-Means [13] vs. C-Means [14]

Nos pontos em seguida (2.2.2 e 2.2.3) serão abordados os trabalhos relacionados onde estas técnicas foram empregues, assim como as características que estas abordagens impactaram o respetivo trabalho.

2.2. Técnicas clássicas de segmentação de imagem

Após uma investigação sobre as técnicas apresentadas anteriormente e os projetos realizados chegou-se à conclusão que poucos seriam os casos que coincidiriam com o tema presente na dissertação, sendo a fonte de imagens usada nestes bastante díspar da fornecida para esta dissertação. Na maior parte dos casos de segmentação de espaços florestais estes teriam como base imagens de satélite, o que leva a que as abordagens necessárias sejam bastante diferentes de imagens fornecidas por torres de videovigilância. Por serem imagens retiradas de um panorama mais raso estas captam mais objetos em comparação com as imagens de satélite, como ventoinhas eólicas e as respetivas pás, ou um edifício inteiro, em vez da identificação apenas do seu telhado. Posto isto, optou-se por analisar os trabalhos existentes de forma separada, com o objetivo de realizar um protótipo com base nas diferentes técnicas à disposição, analisando as vantagens e desvantagens que cada técnica teria em função das imagens a segmentar nesta dissertação.

2.2.1. Segmentação por contornos

Na identificação de um objeto numa imagem, a deteção dos contornos é uma das primeiras técnicas a considerar numa segmentação. Numa imagem os contornos são uma das partes mais essenciais, pois estes estão contidos em diversas características bastante relevantes numa imagem, podendo afirmar que são os contornos que fazem uma imagem, contendo cantos, linhas, curvas, etc. A deteção por contornos diminui a quantidade de informação a ser processada, remove a informação desnecessária e armazena a informação essencial para a estrutura de uma imagem [5]. Existindo ainda diversos tipos de técnicas de deteção de contornos, como representado na Figura 6.

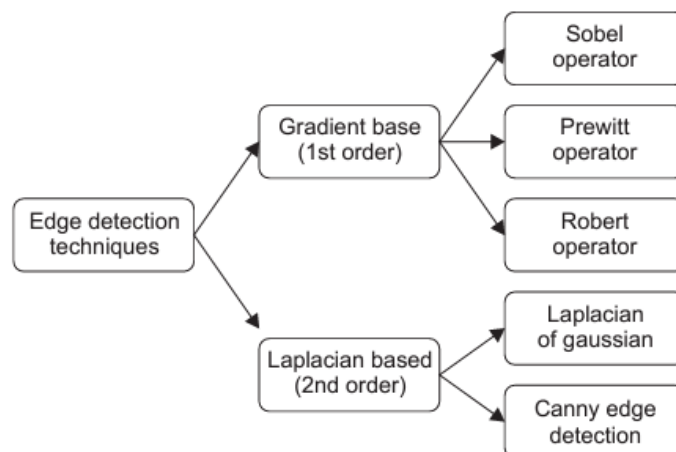


Figura 6: Técnicas de segmentação por contornos [5]

A segmentação por contornos é muitas vezes aplicada em imagens médicas, que servem de auxílio, por exemplo, na identificação de tumores cerebrais, como ilustrado na Figura 7. Esta aplicação que é referida em [7], onde é realizado um processo de comparação entre as diferentes abordagens na segmentação por contornos, concluindo-se que a abordagem *Canny Edge Detection* é a abordagem que melhor se adequa a imagens de alta resolução. No entanto, os resultados obtidos com esta técnica são muito sensíveis ao ruído.

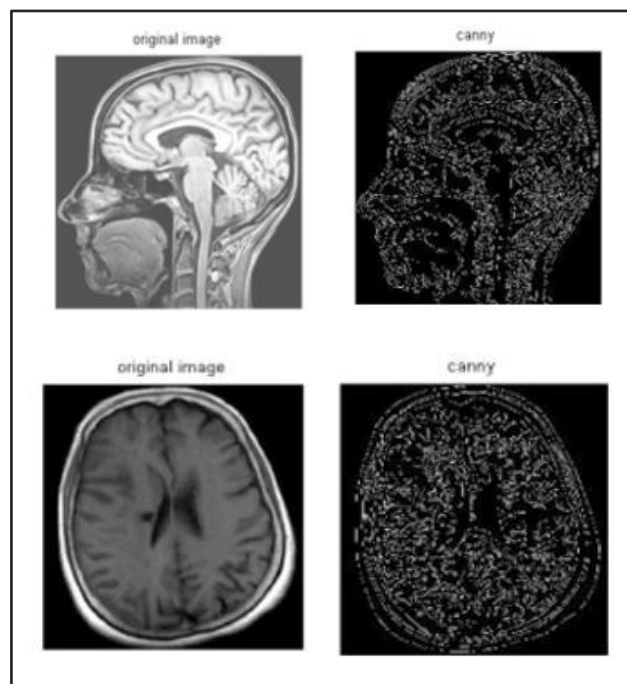


Figura 7: Segmentação por contornos, Canny Edge Detection [7]

Após a análise das técnicas da segmentação por contornos, conclui-se que este género de técnicas não seria útil ao protótipo em questão, pois está-se perante imagens com elevada informação, contornos esbatidos e com ruído.

2.2.2. Segmentação por região

Enquanto uma técnica baseada nos contornos de uma imagem explora os limites dos objetos existentes numa imagem e depois é descoberto o próprio objeto, uma técnica por região adota a abordagem oposta. Nesta abordagem, é atribuído a cada pixel um determinado objeto ou região, dividindo uma imagem em regiões que estão próximas de acordo com um conjunto de critérios pré-definidos [15][16]. Sendo uma forma de segmentar uma imagem com base nos valores dos pixels existentes, esta começa, por exemplo, pelo interior de um objeto e depois vai aumentando a sua área para fora até

encontrar os limites do objeto. As técnicas de segmentação por região são principalmente usadas na identificação de regiões com atributos semelhantes numa imagem. Incluem principalmente dois métodos: Região de Crescimento e Divisão e Fusão da Região [17].

Na Figura 8 é possível observar a aplicação de um desses métodos, o método de Região por Crescimento, a imagens de satélite, onde o objetivo seria segmentar os edifícios existentes nas imagens.

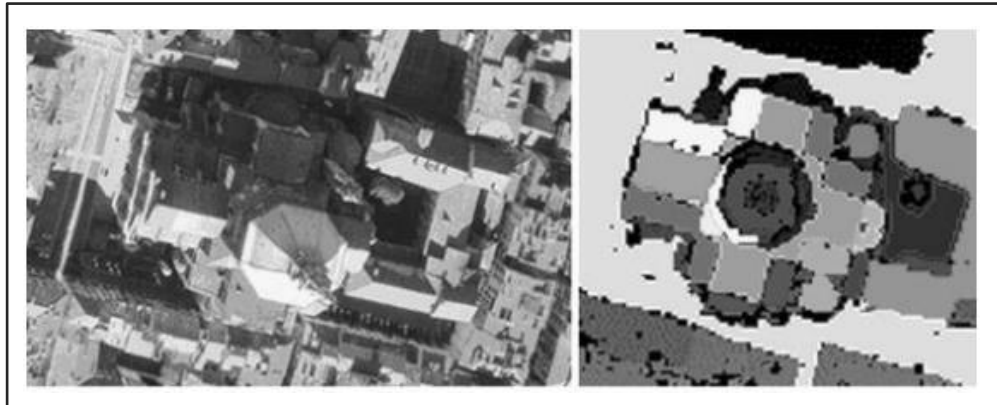


Figura 8: Implementação da técnica de segmentação por região [18]

2.2.3. Segmentação por *thresholding*

A segmentação por *thresholding* é um dos métodos mais utilizados na segmentação de imagem, sendo principalmente usada na identificação de objetos num plano de fundo. Ao selecionar um *threshold*, uma imagem em tons de cinza pode ser convertida numa imagem binária [19][20][21]. Na análise do trabalho proposto em [8], é possível observar a aplicação desta técnica de segmentação a diversos espaços florestais e zonas citadinas, Figura 9 e Figura 10.

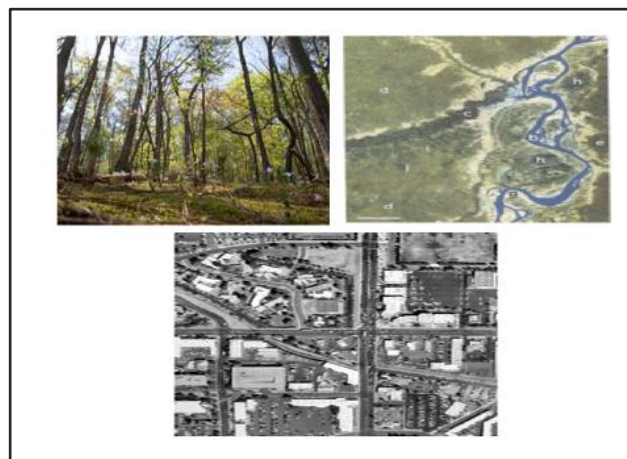


Figura 9: Imagens de zona de floresta e zona citadina [8]



Figura 10: Segmentação de zona de floresta e zona cidadinas [8]

Para além da aplicação desta técnica em imagens em tons de cinza com vista a uma classificação binária de uma imagem, esta também pode ser aplicada em espaços de cor. Para tal é habitual a conversão do espaço de cor RGB (*Red, Green, Blue*) para HSV (*Hue, Saturation, Value*), sendo a segmentação da imagem realizada neste espaço de cor [8]. Na Figura 11 podemos observar a aplicação desta técnica, exibindo os diferentes objetos que poderão ser segmentados. Através desta escala e estabelecendo os valores inferiores e superiores para o objeto a segmentar desejado é possível a segmentação de uma imagem.

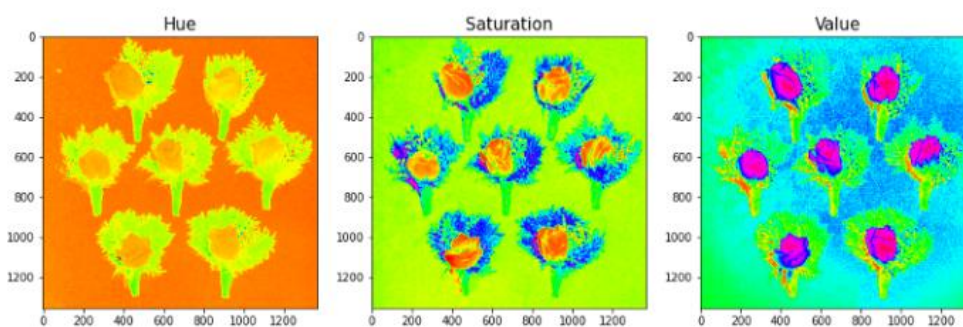


Figura 11: Exemplo de imagens no espaço de cor HSV [12].

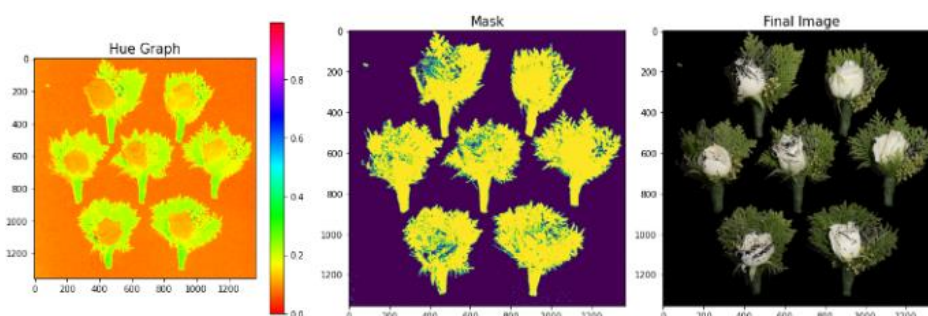


Figura 12: Exemplo de imagens segmentadas [12]

Verificou-se com a aplicação desta técnica de segmentação uma distinção bem-sucedida dos espaços florestais em relação a um segundo plano, como se pode observar na Figura 12. No entanto esta técnica de segmentação apenas funciona corretamente em imagens em que as zonas a segmentar tenham um conjunto de fatores diferenciados da zona de segundo plano, algo que não se evidencia nos dados desta dissertação, onde muitas das vezes a cor existente na zona urbana difunde-se com a zona de floresta.

2.3. Técnicas de aprendizagem automática de segmentação de imagem

2.3.1. K-Means

A segmentação por *clustering* é um tipo de algoritmo de aprendizagem não supervisionado, que é bastante utilizado na segmentação de imagens, sendo o método K-Means um dos algoritmos mais utilizados quando se pretende elaborar a segmentação de uma imagem [22]. Um desses casos é o representado no trabalho proposto em [23], onde foi utilizado este algoritmo para a segmentação de imagens de leucemia. Na detecção da leucemia é importante a segmentação dos glóbulos sanguíneos, e para isso foi utilizado o algoritmo K-Means, que exibiu resultados bastante satisfatórios (Figura 13). A diferença entre os leucócitos (glóbulos brancos) e o fundo das imagens é facilmente visível.

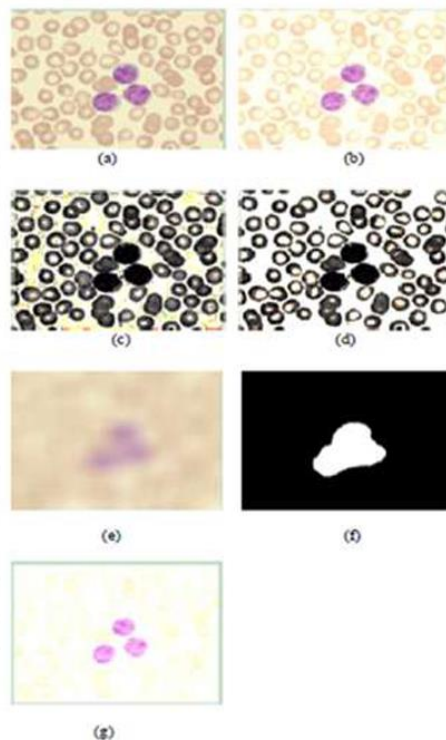


Figura 13: Aplicação do algoritmo K-Means a imagens de leucemia [16]

2.3.2. Redes Neurais Convolucionais

Muitas são as arquiteturas de RNCs existentes, e para uma melhor compreensão da eficácia de cada uma é importante haver uma comparação destas focada numa única base de dados. Com base no trabalho em [25] podemos realizar essa comparação sobre a base de dados Cityscapes [26].

Cityscapes é uma base de dados de grande escala, que contém anotações com um alto nível de detalhe referente a 5000 imagens recolhidas em cenas de rua de 50 cidades diferentes. Seguindo o protocolo de avaliação [18], são utilizadas para avaliação 19 etiquetas semânticas pertencentes a 7 categorias: solo, construção, objetos, natureza, céu, humanos e veículos. Os conjuntos de treino, validação e teste contêm 2975, 500, e 1525 imagens, respetivamente.

Esta base de dados torna-se interessante de analisar devido ao semelhante nível de detalhe e informação existente em comparação com a base de dados utilizada nesta dissertação. Podendo assim analisar os resultados já obtidos nas diferentes abordagens, e realizar posteriormente uma filtragem destas com vista à aplicação na dissertação.

Para este estudo foram escolhidas um total de 18 redes neurais convolucionais: DeepLabv2-CRF [25], DeepLayerCascade [27], ML-CRNN [28], Adelaidecontext [29], FRRN [30], LRR-4x [31], RefineNet [32], FoveaNet [33], Ladder Dense Net [34], PEARL [35], Global-Local-Refinement [36] SACmultiple [37], SegModel [38], TuSimpleCoarse [39], Netwarp [40], ResNet-38 [41], PSPNet [42] e DeepLabv3[25]. De forma a saber qual seria a arquitetura com melhores resultados foi aplicada o método de avaliação IOU (Intersection Over Union) [42].

Ao observar a Tabela 1, referente à performance das diversas arquiteturas abordadas, pode afirmar-se que a arquitetura DeepLabv3 é a que apresenta melhores resultados na segmentação das imagens existentes na base de dados em questão. Detendo o valor médio IOU mais elevado em comparação com as restantes arquiteturas, esta abordagem poderá ser uma forte candidata à aplicação ao protótipo a ser construído.

Método	mIOU*
DeepLabv2-CRF [25]	70,4
DeepLayerCascade [27]	71,1
ML-CRNN [28]	71,2
Adelaidecontext [29]	71,6
FRRN [30]	71,8
LRR-4x [31]	71,8
RefineNet [32]	73,6
FoveaNet [33]	74,1
Ladder Dense Net [34]	74,3
PEARL [35]	75,4
Global-Local-Refinement [36]	77,3
SACmultiple [37]	78,1
SegModel [38]	79,2
TuSimpleCoarse [39]	80,1
Netwarp [40]	80,5
ResNet-38 [41]	80,6
PSPNet [42]	81,2
DeepLabv3 [25]	81,3

*média Intersection-Over-Union

Tabela 1: Performance de diversos métodos sobre a base de dados Cityscapes [17].

A arquitetura DeepLabv3 é uma arquitetura desenvolvida pela Google, correspondendo à terceira versão da arquitetura DeepLab. De forma a lidar com o problema da segmentação de objetos em diferentes planos, esta arquitetura aplica uma *atrous convolution* (expande o tamanho da janela sem aumentar o seu peso, inserindo valores zero em núcleos de convolução [10]), em cascata ou em paralelo de forma a captar todos os objetos nas múltiplas escalas [13].

Esta arquitetura utiliza um módulo ASPP (Atrous Spatial Pyramid Pooling) melhorado, incluindo a normalização de *batches* e características a nível de imagem.

O modelo DeepLabV3 segue a seguinte arquitetura:

- As características são extraídas da rede de base (VGG, DenseNet, ResNet).
- Para controlar a dimensão do mapa de características, é utilizada a *atrous convolution* nos últimos blocos.
- Para além das características extraídas, é adicionada uma rede ASPP para classificar cada pixel correspondente às suas classes.

- A saída da rede ASPP é passada por uma convolução de 1 x 1 para obter o tamanho real da imagem, que será a máscara segmentada final para a imagem.

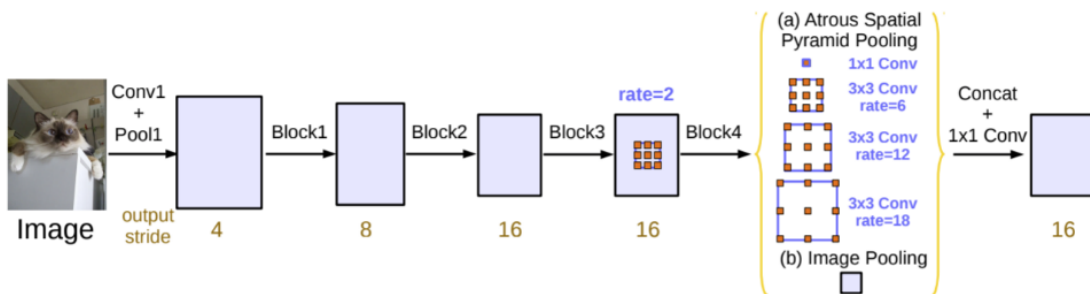


Figura 14: Processo da arquitetura DeepLabv3 [43]

É possível observar na Figura 15 alguns exemplos de segmentações resultantes da aplicação desta arquitetura, encontrando-se os diferentes elementos existentes na imagem devidamente segmentados.

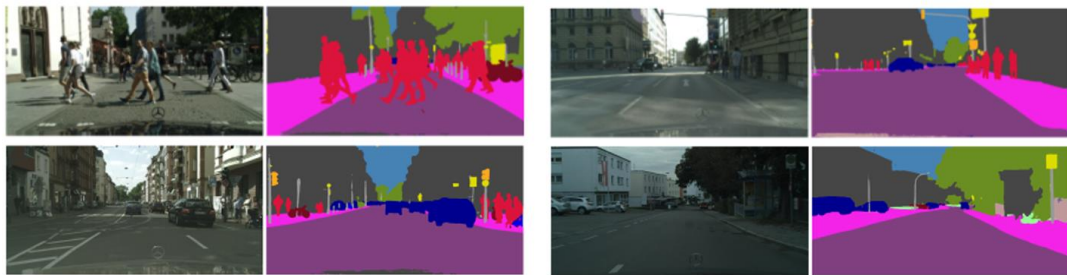


Figura 15: Segmentação de imagem através do método DeepLabV3 [24]

2.4. Processos de tratamento de imagem

Na maioria dos processos de segmentação de uma imagem é necessário realizar algumas transformações às mesmas, seja com vista à obtenção de resultados diferentes, ou tendo em vista a redução do ruído existente, pois este introduz dificuldades no processo de segmentação [44]. Com o objetivo da remoção desse ruído, como por exemplo o nevoeiro existente em imagens, em [45] é apresentado um processo de remoção efetivo através do uso do *dark channel prior*.

O *dark channel prior* (DCP) baseia-se na propriedade de "pixéis escuros", que têm uma intensidade muito baixa em pelo menos um canal de cor, exceto na região do céu. Devido à sua eficácia na remoção do nevoeiro, a maioria das técnicas de remoção de nevoeiro adotaram o DCP. Estas técnicas baseadas no DCP são compostas por quatro

etapas principais: estimativa da luz atmosférica, estimativa do mapa de transmissão, apuramento do mapa de transmissão, e reconstrução da imagem [44].

Como se pode observar na Figura 16, este método torna-se bastante eficaz em imagens totalmente preenchidas com nevoeiro, facilitando a segmentação de uma imagem, que poderia ficar comprometida.



Figura 16: Exemplo antes e depois da utilização do algoritmo de remoção de nevoeiro [45]

Análise e Exploração dos Dados

Ao longo deste capítulo serão exploradas as imagens provenientes de torres de vigilância, com vista à elaboração dos modelos de visão por computador propostos, com o objetivo de segmentar espaços florestais de espaços urbanos, referindo quais as bases de dados escolhidas para serem utilizadas para cada modelo.

3.1. Conjunto de Dados

Para a realização desta dissertação foram fornecidas um total de 128 imagens referentes a espaços florestais, acompanhadas das suas máscaras respectivas, perfazendo um total de 256 imagens.

Sendo a amostra de dados insuficiente para a construção de uma rede neuronal convolucional fidedigna houve assim a necessidade de realizar a divisão de cada imagem e respectivas máscaras em 9 sub-imagens, resultando desta forma um total de 2560 imagens, sendo este total constituído pela imagem original e as 9 sub-imagens correspondentes.

Existiu ainda a necessidade da elaboração de uma base de dados em formato csv. de forma a identificar os espaços florestais abordados e o número de paisagens em cada área. Facilitando assim uma análise das imagens e posterior avaliação dos resultados obtidos para os diferentes espaços florestais, seguindo o formato da Tabela 2. A Tabela 2 é constituída por 6 campos, respetivamente, id, nome da imagem, diretoria da pasta onde a imagem está armazenada, diretoria da imagem, área correspondente e diretoria da máscara respetiva.

Id	image	folder_path	image_path	area	path_mask
AREF_12_FSJ...	AREF_12_FSJ...	C:\Users\dafdi...	C:\Users\dafdi\...	FSJ	C:\Users\dafdi\...
AREF_12_FSJ...	AREF_12_FSJ...	C:\Users\dafdi...	C:\Users\dafdi\...	FSJ	C:\Users\dafdi\...
AREF_12_FSJ...	AREF_12_FSJ...	C:\Users\dafdi...	C:\Users\dafdi\...	FSJ	C:\Users\dafdi\...
AREF_12_FSJ...	AREF_12_FSJ...	C:\Users\dafdi...	C:\Users\dafdi\...	FSJ	C:\Users\dafdi\...
AREF_12_FSJ...	AREF_12_MAT...	C:\Users\dafdi...	C:\Users\dafdi\...	MATOU	C:\Users\dafdi\...
AREF_12_MAT...	AREF_12_MAT...	C:\Users\dafdi...	C:\Users\dafdi\...	MATOU	C:\Users\dafdi\...
AREF_12_MAT...	AREF_12_MAT...	C:\Users\dafdi...	C:\Users\dafdi\...	MATOU	C:\Users\dafdi\...
AREF_12_MAT...	AREF_12_MAT...	C:\Users\dafdi...	C:\Users\dafdi\...	MATOU	C:\Users\dafdi\...
...					

Tabela 2: Tabela de dados.

Após a exploração das imagens disponibilizadas podemos identificar 6 espaços florestais denominados FSJ, Marinha Grande, Sico, Alvaro, FerrariaSJoão e MATOU correspondendo a cada uma 10, 450, 280, 280, 250 e 10 imagens, respetivamente, como se pode observar na Figura 17.

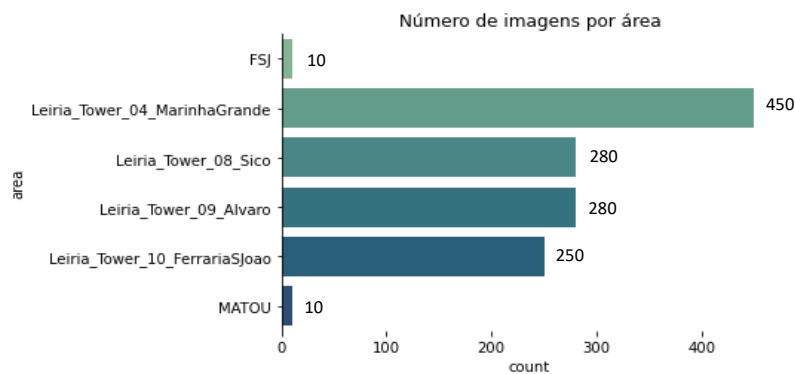


Figura 17: Distribuição das imagens por zonas de florestação

Cada imagem referida como “original“ corresponde à imagem fornecida pelo INOV e tem uma dimensão de 1920 x 1080, enquanto as sub-imagens realizadas sobre as imagens originais têm uma dimensão de 640 x 360. Uma vez que cada grupo de imagens (original e sub-imagens) será tratado separadamente, não existiu a necessidade de proceder a uma normalização da base das imagens.

Como referido anteriormente as imagens poderão apresentar algum ruído, sendo este um dos principais fatores a ter em conta no processo de tratamento dos mesmos. Outro fator a considerar é a dificuldade em identificar a linha do horizonte em algumas imagens, problema este que pode ser causado pelo ruído existente ou a distância a que esta se encontra. Este fator pode dificultar a segmentação, uma vez que não é possível identificar onde termina o espaço florestal.

Em relação às sub-imagens criadas, estas apresentarão uma falta de informação devido a serem um excerto da imagem original, perdendo alguns elementos que faziam parte da imagem original. Quanto às máscaras, e dado que foram construídas manualmente, apresentam algumas imperfeições, como alguma falta de exatidão nas áreas assinaladas como zonas urbanas. No entanto, esta questão não deverá ser um obstáculo na realização do protótipo a construir segundo técnicas mais clássicas e por agrupamento (*clustering*), pois não se baseiam nas máscaras previamente construídas.

A Figura 18 representa uma amostra das imagens originais acompanhadas das máscaras construídas manualmente. Já na Figura 19 é possível observar uma das sub-imagens que corresponde a uma das imagens originais representadas, assim como a respetiva máscara.



Figura 18: Imagens e respetivas máscaras fornecidas pelo INOV



Figura 19: Subimagens realizadas às imagens e máscaras fornecidas

3.2. Tratamento do conjunto de dados

Como referido anteriormente, após a análise das imagens, verificou-se algum ruído, como por exemplo a presença de nevoeiro em diversas imagens, existindo assim a necessidade de tratamento das imagens. De forma a corrigir esta questão, foi realizada uma remoção do nevoeiro recorrendo ao algoritmo “dehazing”, adaptado com base no artigo referido no Capítulo 2 [46], onde é demonstrado o seu processo de tratamento de imagem, destacando que este

tratamento foi implementado em toda a base de dados, uma vez que não se verificou danos em imagens em que não seria necessária a sua implementação.

Após a aplicação deste algoritmo é possível verificar uma redução considerável do ruído existente nas imagens, como observado na Figura 20. Desta forma, é aparentemente ultrapassado um dos fatores problemáticos à aplicação do modelo proposto.



Figura 20: Tratamento da imagem quanto ao ruído

Desenvolvimento

Como visto no capítulo anterior, diversos são os métodos existentes para a realização da segmentação de uma imagem, seja através da cor, contornos, luminância, divisão por *clusters* ou redes neuronais convolucionais. Com isso, houve a necessidade de escolher quais os métodos de segmentação de imagem que seriam abordados dentro do estudo realizado anteriormente, indo de acordo com as características de segmentação e necessidades existentes.

Após a análise das imagens fornecidas foi descartado logo à partida um dos métodos clássicos de segmentação, a segmentação por contornos. Devido ao elevado nível de detalhe das imagens tornou-se impossível a realização de contornos das zonas urbanas, sendo este o principal motivo de exclusão. Outro método de segmentação clássico seria a segmentação por *thresholding*. É possível deduzir que certas zonas urbanas se confundam com espaços florestais devido à projeção da luz solar nas diversas alturas do dia, ou a existência de zonas de habitação mais escuras que irão dificultar a definição de um *threshold*, no entanto, este método será analisado posteriormente como possível solução.

Sobram então os métodos por divisão de *clusters*, *K-Means*, e a utilização de uma rede neuronal convolucional. Não existindo à partida alguma contrariedade, estes serão os outros dois métodos em análise, sendo retratado neste capítulo o detalhe de cada processo de segmentação.

Em concordância com o INOV foi estabelecido que seria preferível que a máscara a ser realizada fosse mais reduzida, podendo deixar zonas urbanas por identificar, mas nunca cobrindo espaços florestais indevidamente. Deste modo, poderiam continuar a existir falsos negativos, possíveis incêndios detetados em zonas urbanas, no entanto, teria de ser garantido que nenhum espaço florestal ficaria coberto, levando a que a um possível incêndio nunca chegasse a vir a ser detetado.

Tendo este fator em consideração, em primeiro lugar será retratado o método mais clássico de segmentação, segmentação por *thresholding*. Em seguida será desenvolvido o método através da divisão por *clusters*, *K-Means*, e por consequente a rede neuronal convolucional DeepLabv3. Sendo estes dois modelos construídos com base nos dados já devidamente tratados quanto ao ruído.

4.1. Thresholding

Apesar dos aspetos negativos referidos anteriormente em relação à técnica de segmentação por *thresholding*, foi testada inicialmente uma abordagem com base na luminância das imagens, onde com base em múltiplas imagens foi definido um *threshold* para cada uma delas, construindo assim uma máscara para um determinado espaço florestal em diversas alturas do dia. Assim, foi possível captar todos os pontos de luminância nas diversas alturas do dia, tentando contornar ao máximo zonas de sombra que poderão ser captadas nas imagens, que iriam tornar uma zona urbana clara, numa zona mais escura, confundido esta com um espaço florestal. Após a elaboração das máscaras para as diversas alturas do dia, estas foram unidas resultando numa única máscara.

O valor do *threshold* definido, é calculado com base na luminância existente na imagem, sendo a solução desenvolvida baseada na média de luminância existente na imagem. Estes parâmetros são configuráveis, tendo sido previamente definidos pelo INOV.

Este método torna-se simples e obsoleto comparado com outras abordagens, contendo apesar disso a vantagem de ser um método prático e não haver uma necessidade de treino.

4.2. K-Means

Para a elaboração deste modelo será utilizado como fonte a base de dados na dimensão original fornecida pelo INOV, desenvolvendo o modelo apenas para uma imagem em particular, comparando posteriormente com a máscara respetiva elaborada manualmente como método de avaliação dos resultados.

Tendo as imagens devidamente tratadas procedeu-se à criação de um algoritmo recorrendo ao modelo *K-Means*. Este algoritmo é suportado pela biblioteca OpenCV, onde através da função *K-Means* é feita a segmentação de cada imagem em particular. Para a realização deste modelo foi preferido este devolver dois *clusters*, uma vez que ao analisar as imagens, não existiria mais detalhe para além de zonas urbanas e espaços florestais. De notar que a zona correspondente ao céu também seria identificada neste algoritmo, no entanto, o resultado obtido para esta seria irrelevante, uma vez que o algoritmo desenvolvido pelo INOV já eliminaria esta zona.

Este algoritmo tem como base o volume do espaço florestal vs. volume de zona urbana, existindo a necessidade de analisar em primeiro lugar a informação da imagem a segmentar, de forma a perceber qual o *cluster* deverá ser removido, seguindo o esquema abaixo.

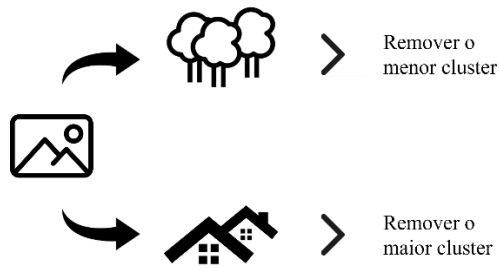


Figura 21: Esquema de remoção de clusters

O esquema representado na Figura 21 representa o *input* de uma única imagem, sendo posteriormente alvo de uma decisão consoante a constituição da imagem, seguindo a seguinte lógica: Caso o *input* recebido pelo algoritmo tenha uma predominância de espaços florestais, o *cluster* a ser removido deverá ser o de menor tamanho, no entanto, se houver uma predominância de zonas urbanas o *cluster* a ser removido deverá ser o de maior tamanho.

Para que este esquema seja bem-sucedido, é imperativo que esta análise seja realizada com sucesso, para isso foi necessário realizar uma exploração *a priori* de cada foto, sendo a predominância de cor clara o critério para a identificação dos *clusters* a eliminar.

Conhecendo o *cluster* a eliminar, a variável *k* inicia aleatoriamente sobre os 2 diferentes *clusters* e uma variável denominada “center” define o centro destes *clusters*. A distância de cada ponto destes centros é calculada e depois são atribuídos a um dos conjuntos. Posteriormente são divididos em diferentes segmentos de acordo com o valor da *label* pré-definida, zona urbana ou espaço florestal.

Após a aplicação do modelo *K-Means* e da divisão dos conjuntos pelas *labels*, proceder-se-á à remoção da *label* desejada.

4.3. DeepLabv3

Na construção do modelo de redes neuronais convolucionais DeepLabv3, e ao contrário do modelo *K-Means*, este recorre ao conjunto total de sub-imagens obtidas anteriormente. Para a construção do modelo existiu a necessidade de dividir os dados em treino, validação e teste, tanto para as imagens de espaços florestais como para as máscaras correspondentes.

4.3.1. Tratamento dos dados

A base de dados constituída por imagens de espaços florestais e máscaras foi separada em 6 conjuntos: treino e treino_labels, constituídos cada um por 1053 imagens, validação e validação_labels, constituídos por 63 imagens cada um, e teste e

teste_labels, que irão permitir avaliar os resultados do modelo, constituídos cada um por 18 imagens.

Já com as imagens devidamente distribuídas pelas pastas, em seguida, foi realizada uma pré configuração das máscaras que serão construídas com os valores RGB que a iriam constituir. Para isso, foi construído um ficheiro .csv onde foram atribuídos os valores RGB (0,0,0) para a categoria “background” e (255,255,255) para a categoria “building”. Através desta configuração as zonas urbanas detetas serão pintadas a preto, enquanto a restante imagem será pintada de branco.

Com a configuração das máscaras a serem obtidas já estabelecida foi realizada uma análise mais aprofundada das imagens e respetivas máscaras. Para isso, foram elaborados dois novos grupos de máscaras. Um constituído por máscaras através do processo One-Hot Encoding (conversão dos dados através de uma avaliação categórica binária), e outro correspondente à transformação das máscaras anteriores para os valores preto e branco, valores estabelecidos anteriormente. Desta forma, foi possível fazer uma categorização das zonas urbanas (a preto) e espaços florestais (a branco), que será necessário no processo de modelação posterior.

4.3.2. Arquitetura do modelo DeepLaV3

A arquitetura do modelo DeepLabV3 teve como base a rede neuronal convolucional ResNet-101. A rede neuronal convolucional ResNet-101 é uma rede neural convolucional com 101 camadas de profundidade, a escolha desta profundidade é justificada pelo detalhe elevado existente nas imagens utilizadas. Sendo que se está perante uma classificação binária das imagens foi a utilizada a função *sigmoid* na camada de *output* da arquitetura do modelo.

4.3.3. Treino e Avaliação

Para a realização dos conjuntos de treino e avaliação foi necessário em primeiro lugar transformar as imagens pertencentes a estes dois conjuntos em *tensors*, de forma a estas poderem ser empregues no modelo construído anteriormente. Devido à dimensão dos conjuntos criados e à rede neuronal convolucional escolhida (Resnet-101), houve a necessidade de optar inicialmente pela escolha de 10 *epochs* (número de vezes que o algoritmo itera sobre o conjunto de dados de treino). Em causa está o tempo que seria despendido para o treino e validação do modelo. Posteriormente este foi treinado com 100 *epochs*, tendo sido utilizado o otimizador Adam em ambos os casos.

Para a realização da avaliação foram utilizados dois métodos de análise, Intersection over Union e Dice Coefficient. O primeiro consiste num rácio, onde como numerador é calculado a área de sobreposição entre a área de delimitação prevista e a área de delimitação real. Já o denominador consiste na área de união, ou seja, a área que engloba tanto a área de delimitação prevista como a área de delimitação real [37].

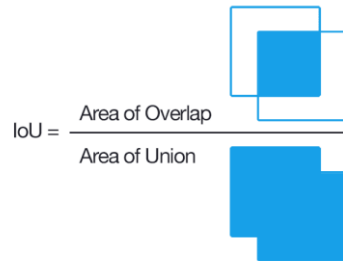


Figura 22: Cálculo do método de avaliação Intersection over Union [37]

O método Dice Coefficient, baseia-se na sobreposição das duas amostras, esta medida varia entre 0 e 1, onde 1 representa uma sobreposição perfeita [37], representado pela expressão em seguida:

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|}$$

Figura 23: Cálculo do método de avaliação Dice Coefficient [17]

$|X \cap Y|$ simboliza os elementos em comum entre X e Y, e $|X|$ e $|Y|$ representam o número de elementos existentes em X e Y, respetivamente.

Avaliação dos Resultados

Este capítulo tem como objetivo a análise e discussão dos resultados dos modelos elaborados, comparando-os, de forma a obter o modelo ideal. Consoante a abordagem utilizada diferentes seriam os resultados previstos, sendo esperado na abordagem por *thresholding* uma máscara completa, em que os pontos de luminância abaixo dos valores estabelecidos ao longo do dia seriam marcados. Do modelo *K-Means* também adviria uma máscara completa, tendo em conta os diferentes espaços RGB, e através dessa identificação seria então realizada a sua segmentação. Já a abordagem recorrida à rede neuronal convolucional, baseada nas máscaras previamente desenvolvidas manualmente, sucederia uma segmentação da imagem original em diversos blocos, uma vez que esta terá sido dividida em 9 sub-imagens.

5.1. Resultados dos modelos

Na análise aos resultados obtidos primeiramente optou-se por realizar uma análise visual das máscaras obtidas. Em todos os processos foram elaboradas diversas máscaras de segmentação, sendo exibidas apenas exemplos que melhor ilustram os resultados obtidos. Após essa análise será realizado uma análise dos modelos de avaliação já referidos, com vista a uma comparação objetiva dos resultados.

5.1.1. Resultados do método por *thresholding*

Ao observar a Figura 24, podemos verificar uma mudança das máscaras realizadas através da segmentação por *thresholding* no decorrer do dia. Posteriormente estas foram moldadas numa só, como referido anteriormente, obtendo a máscara final retratada na Figura 25.

Verifica-se que zonas como telhados não são captadas pelas máscaras através deste método, o que vai de encontro ao que foi previsto anteriormente. Estas, sendo mais escuras não são alcançadas pelos valores de *thresholding* estabelecidos. É também possível observar a existência de falhas na máscara, algo que tentou ser resolvido com as abordagens *erode* e *dilate* da biblioteca Opencv, pela ordem descrita. O primeiro método é usado para executar a erosão de uma imagem, diminuindo assim as características da mesma. O segundo, tem como objetivo aumentar a área de um objeto, acentuando assim as suas características. No entanto, estas tornaram-se insuficientes

aquando deste preenchimento, não cobrindo os espaços necessários, ou por outro lado cobrindo espaços florestais indesejados.



Figura 24: Máscara por thresholding em diversas alturas do dia

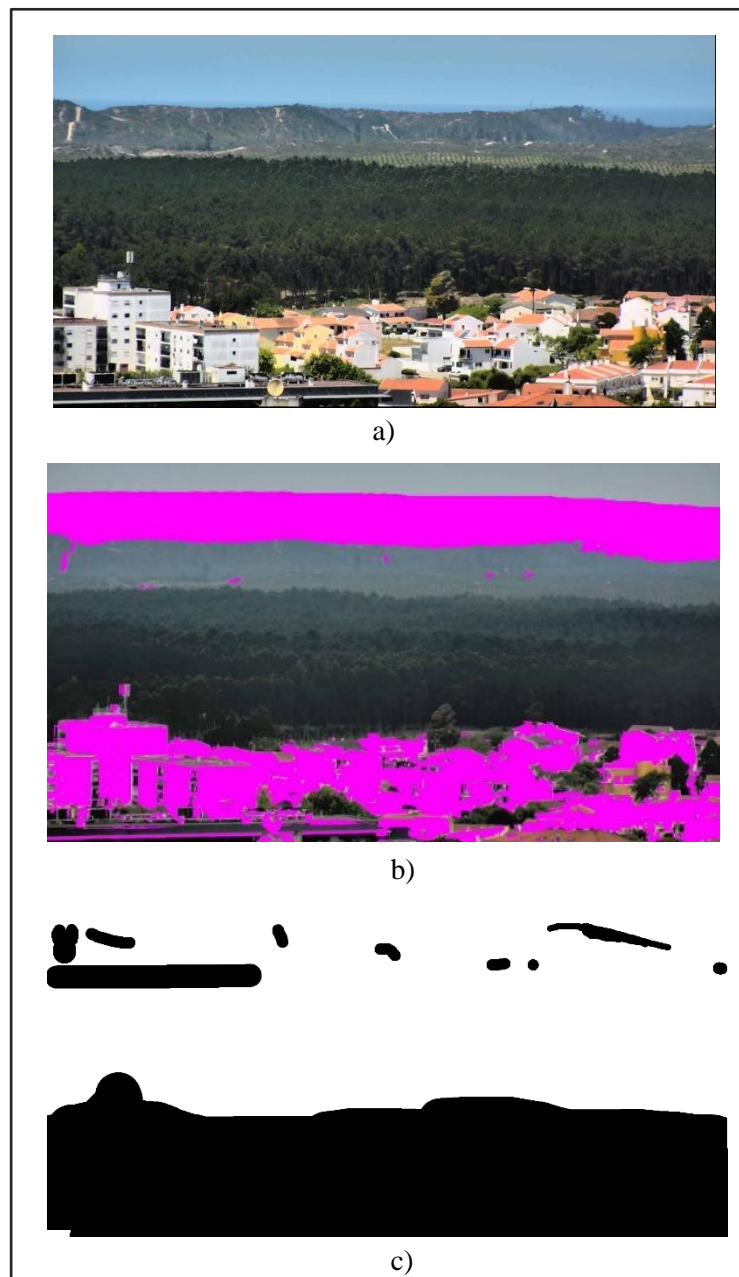


Figura 25: Máscara por Thresholding, a) imagem original, b) máscara criada, c) máscara manual

Para esta abordagem ser corretamente realizada, como foi possível observar, são necessárias diversas imagens dos espaços florestais em diferentes momentos do dia, resultando numa máscara final. Sendo esta o objeto de avaliação final, comparando com a máscara desenhada manualmente. Após aplicados os métodos de avaliação abordados nesta dissertação, Intersection over Union e Dice Coefficient, foi possível chegar à média dos resultados 0.4 e 0.35, respetivamente. Estes resultados são relativamente inferiores, no entanto, algo expetável devido à especificidade dos critérios desta abordagem. O principal motivo é a falta de homogeneidade entre as imagens e na informação existente dentro de cada uma, levando a que a definição de um *threshold* seja bastante vaga quando numa imagem a luminância pode variar.

5.1.2. Resultados do modelo *K-Means*

Ao analisar os resultados obtidos pelo modelo *K-Means*, verifica-se que este obteve excelentes resultados em imagens de alta qualidade, sem ruído, e com um baixo nível de informação, imagens em que os planos de zona urbana e espaços florestais fossem facilmente distinguidos, permitindo realizar uma segmentação de forma facilitada. Verificou-se uma cobertura completa das zonas urbanas desejadas, assim como da zona de céu existente, algo que não seria o objetivo inicial, mas que seria uma benesse na criação das máscaras, uma vez que assim só se contaria com uma imagem onde existiria apenas o espaço florestal.

No entanto nem todos os resultados obtidos foram satisfatórios. Em imagens com diversos planos e muita informação, o modelo acabou por confundir espaços florestais com zonas urbanas, por vezes não cobrindo a totalidade das áreas desejadas, ou cobrindo estas em demasia.

É possível observar na Figura 26 os resultados positivos do modelo *K-Means* quando itera sobre imagens com pouca informação, ou onde as zonas a segmentar são distintas. O mesmo já não ocorre quando se está perante uma imagem com ruído, onde mesmo com o tratamento de remoção executado viu-se incompetente face a certas imagens em que as zonas urbanas e espaços florestais são imprecisos (Figura 27), prolongando em demasia a zona urbana.

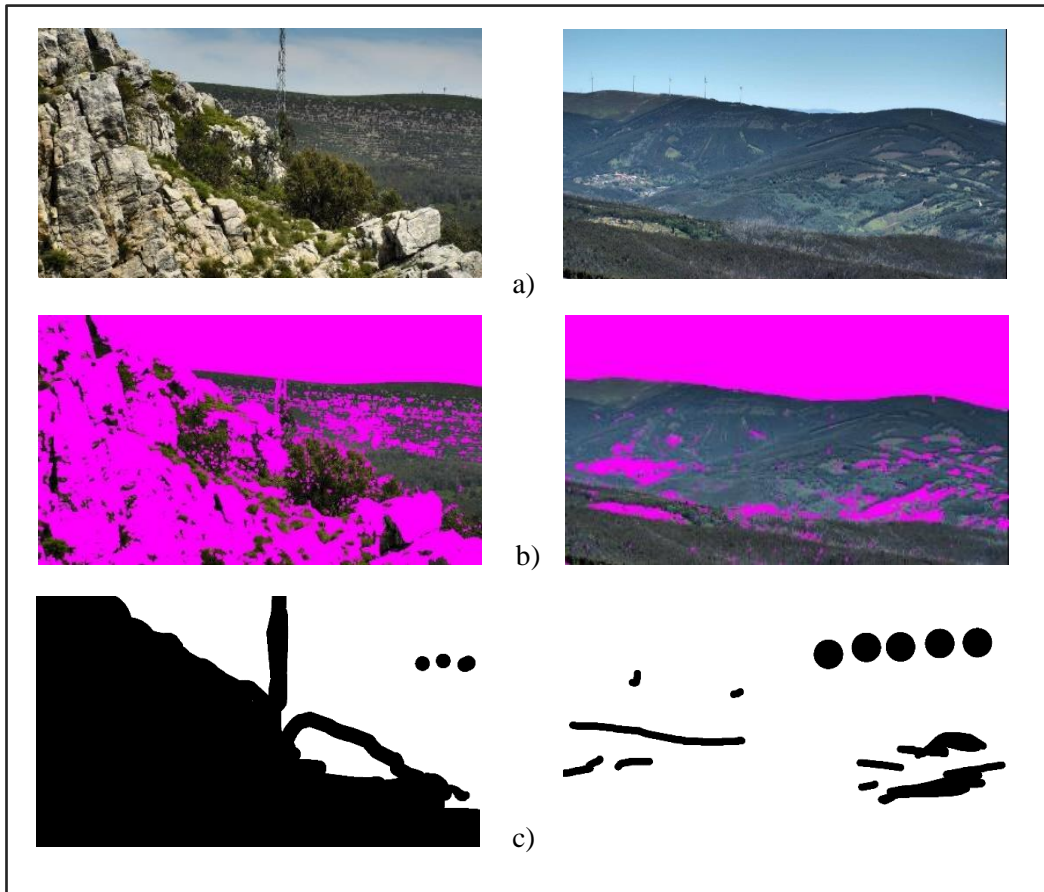


Figura 26: Resultados positivos K-Means, a) imagem original, b) máscara criada, c) máscara manual

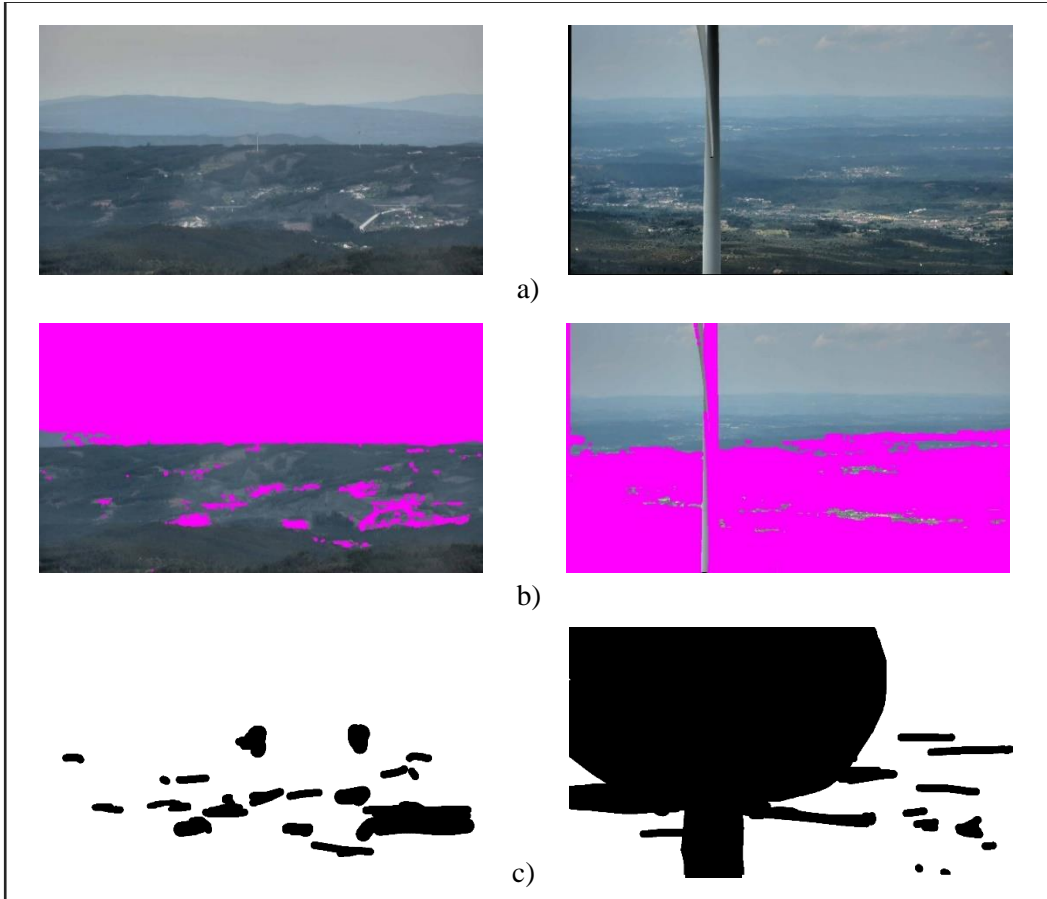


Figura 27: Resultados negativos K-Means, a) imagem original, b) máscara criada, c) máscara manual

A abordagem pelo modelo K-Means apresentou melhores resultados comparativamente à abordagem por *thresholding*, Tabela 3. De notar que esta avaliação foi realizada com base numa seleção de imagens, onde foi selecionada manualmente a segmentação em que o *cluster* desejado foi eliminado, referido no ponto 4.2.

	Intersection over Union	Dice Coefficient
Thresholding	0.40	0.35
K-Means	0.55	0.47

Tabela 3: Comparação de resultados

5.1.3. Resultados do modelo DeepLabV3

Os problemas de detalhe e elevada informação de imagens já não existiram no modelo convolucional, uma vez que este é baseado nas máscaras respetivas a cada imagem. Concluiu-se, com a análise dos *outputs*, que os resultados não aparentavam ser tão satisfatórios como a aplicação do modelo *K-Means*, no entanto, também os erros não seriam tão elevados. Indo de encontro com a proposta realizada pelo INOV, onde seria preferível uma máscara que não preenchesse em demasia uma imagem, podendo assim ser retocada caso fosse identificado algum incidente numa área que deveria estar marcada como zona urbana, não impedindo a não deteção de um incêndio.

Na análise às máscaras finais demonstradas na Figura 28 o modelo evidenciou ir de acordo ao desejado, sendo coberta parte das zonas urbanas desejadas, e não cobrindo em demasia espaços florestais que comprometam a não deteção de incêndios. Pode então concluir-se que o modelo recorrendo à arquitetura DeepLabV3 será o mais adequado na aplicação deste problema, sendo ainda necessário a realização de uma avaliação do modelo construído, de forma a corroborar as conclusões obtidas.

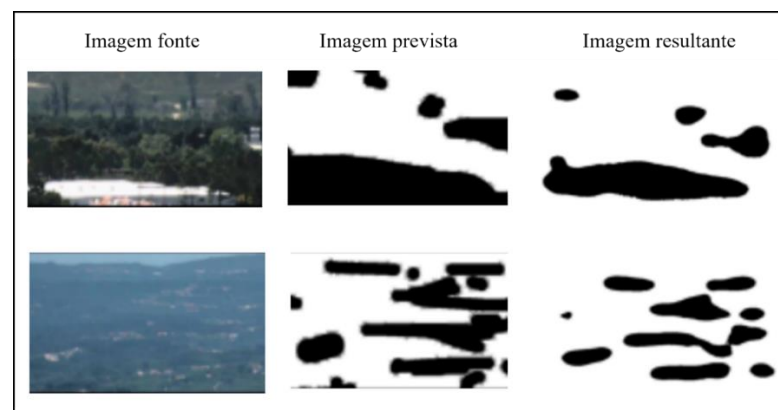


Figura 28: Resultados DeepLabV3

5.2. Avaliação dos resultados

Como descrito anteriormente a rede neuronal convolucional construída com base na arquitetura DeepLabv3 foi a selecionada para o desenvolvimento deste protótipo. Neste sentido será efetuada uma avaliação dos resultados obtidos, de forma a confirmar a escolha do modelo selecionado. Para isso recorreu-se às avaliações Intersection over Union e Dice Coefficient, descritas no capítulo anterior. Na Figura 29 são demonstrados os valores de teste para as duas avaliações referidas, IOU e Dice, obtidos *vs.* o número de *epochs* durante o treino. A linha verde representa o resultado obtido pelo modelo, e os pontos a vermelho o resultado obtido pelo modelo de avaliação. Pode-se verificar que os resultados não são tão favoráveis como esperado, o que pode ser devido à falta de detalhe e qualidade da informação existente nas sub-imagens selecionadas para o modelo, existindo uma larga distância entre os valores de validação e de treino nas duas avaliações, contrapondo com a análise realizada às máscaras produzidas.

A primeira razão que pode motivar estes resultados é a constituição da amostra após a divisão das imagens originais. Sendo divididas em pequenas imagens algumas destas apresentam apenas zonas de céu e/ou zonas florestais, levando a que as máscaras respetivas não tenham nenhuma segmentação desenhada. De forma a contornar este problema existiu a necessidade de executar novamente o modelo construído para duas amostras. Numa primeira tentativa seriam empregues as imagens e máscaras com o tamanho original. A razão desta substituição é a informação existente nas imagens originais comparadas com as sub-imagens retiradas destas. Por ser uma sub-imagem o nível de informação retirado no treino do modelo seria insatisfatório para a obtenção de resultados fidedignos, não existindo assim uma amostra desbalanceada. De assinalar que a utilização destas imagens leva também a um treino mais débil, derivado da escassa dimensão da amostra. Numa outra abordagem foram removidas as imagens e respetivas máscaras que não seriam de interesse para o modelo, e estariam a prejudicar a segmentação deste, imagens com as características referidas anteriormente. Nesta abordagem os conjuntos de validação, treino e teste passaram a ser constituídos por 50, 756 e 13 imagens, respetivamente.

Após uma nova avaliação do modelo à abordagem com as imagens originais podemos observar na Figura 30 que este obteve melhores resultados que os anteriores. No entanto não apresentam também ser muito satisfatórios, demonstrando um padrão dos pontos obtidos pelos métodos de avaliação bastante aleatório nos dois gráficos, refletindo uma falta de acerto do modelo, o que demonstra que não foi executada uma segmentação das imagens com sucesso. Na segunda abordagem o modelo apresentou o melhor valor médio de Dice Coefficient, 0,78,

já a avaliação Intersection over Union apresentou o valor médio 0,51, Tabela 4. Pode-se concluir que a causa do problema existente na primeira avaliação seria o desbalanceamento da amostra, apesar do baixo valor no segundo método de avaliação em comparação ao primeiro.

Para além do desbalanceamento da amostra outros motivos poderão estar por trás do fraco desempenho. Um dos motivos a apontar é o facto das máscaras utilizadas não marcarem toda a zona urbana de forma rigorosa (por serem elaboradas manualmente), levando assim a um fraco treino do modelo, e por consequente uma fraca performance do mesmo. Outro motivo para a causa destes resultados é a utilização dos métodos de avaliação que se baseiam na sobreposição das máscaras criadas automaticamente sobre as máscaras realizadas manualmente. Sendo que estas apenas cobrem em parte as zonas desejadas leva a que em comparação com as máscaras elaboradas manualmente seja recorrente haver pontos de não coincidência.

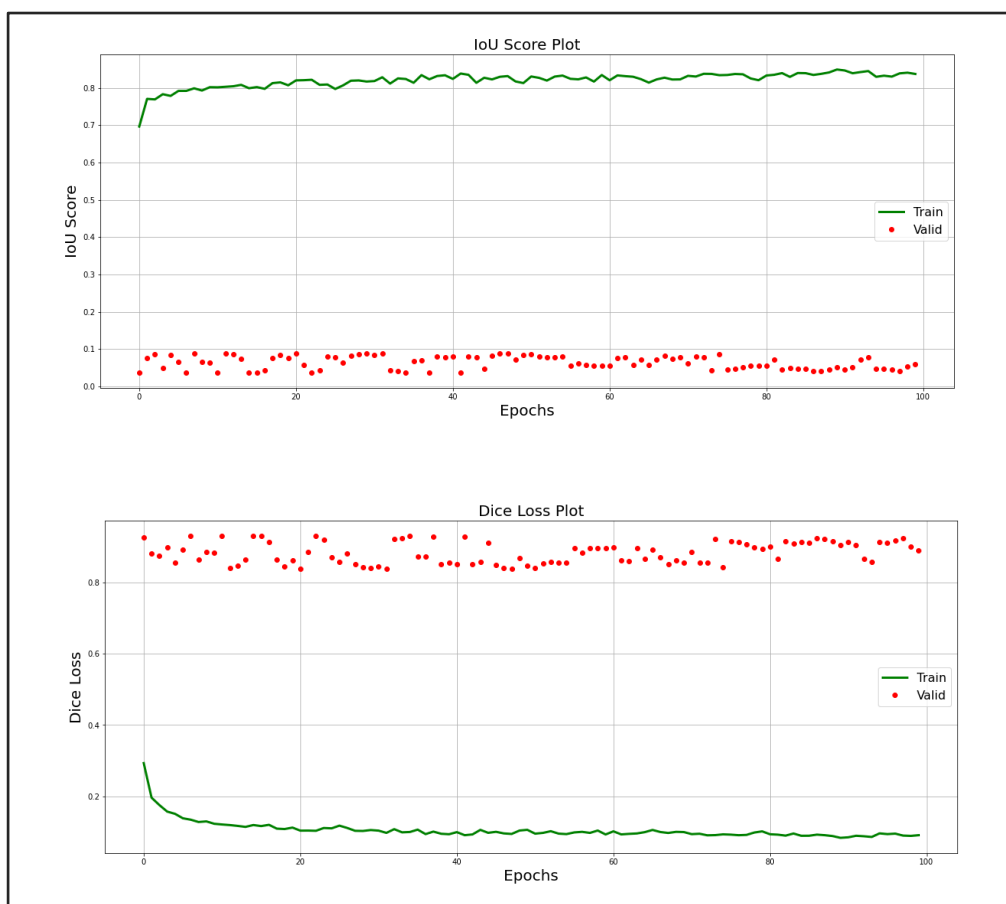


Figura 29: Avaliação do modelo DeepLabV3 recorrendo às subimagens realizadas

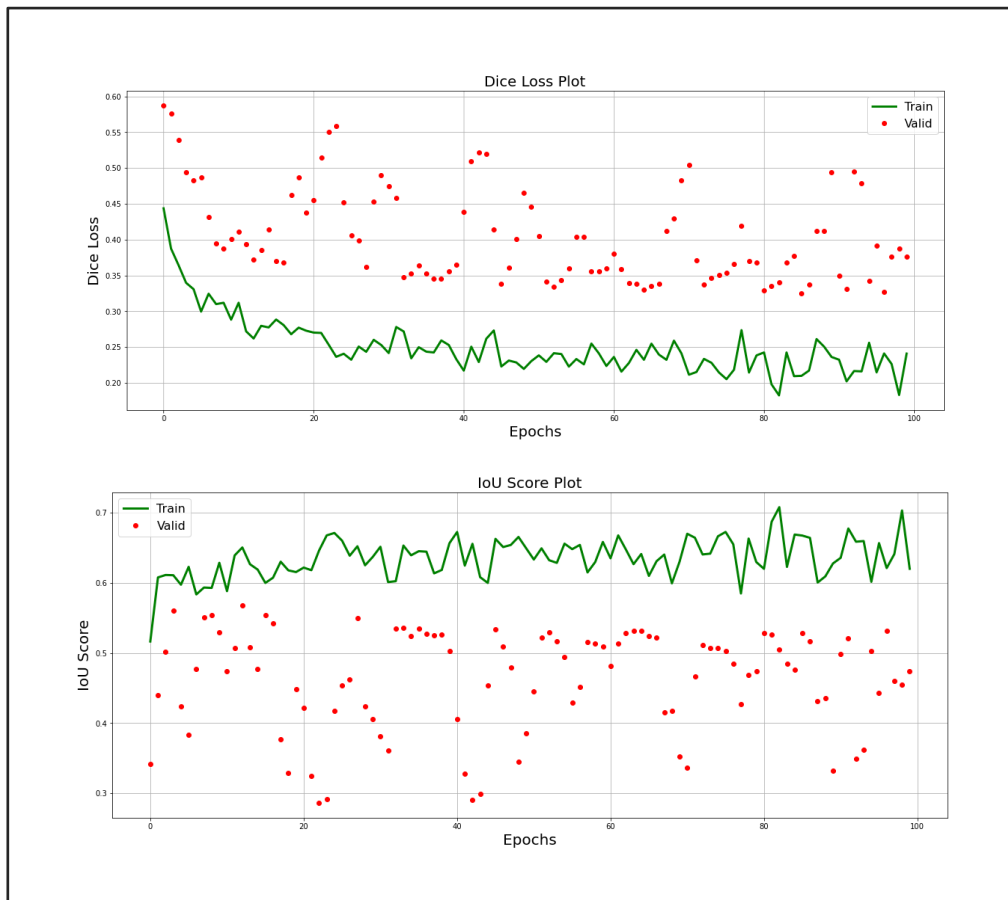


Figura 30: Avaliação do modelo DeepLabV3 recorrendo às imagens com tamanho original

	Intersection over Union	Dice Coefficient
Thresholding	0,40	0,35
K-Means	0,55	0,47
DeepLabv3	0,51	0,78

Tabela 4: Resultados dos métodos de avaliação

CAPÍTULO 6

Conclusões

Este projeto de investigação visou encontrar uma solução alternativa automatizada para a segmentação das imagens utilizadas pelo INOV no combate aos incêndios florestais, uma vez que são elaboradas manualmente. O objetivo foi uma redução da mão de obra necessária e redução dos erros humanos existentes, traduzindo-se num combate mais eficaz aos incêndios florestais. Com as imagens fornecidas pelas torres de videovigilância instaladas pelo INOV foi iniciado um processo de investigação de forma a descobrir a melhor forma de automatização para a segmentação das zonas urbanas em espaços florestais, e em que não houvesse a necessidade de intervenção humana. Tendo em consideração que seria preferível uma segmentação menos completa, que não cobrisse espaços florestais potencialmente favoráveis a incêndios, e que ao mesmo tempo não cobrisse também toda a zona urbana, a uma segmentação que cobrisse toda a zona urbana e não garantisse que todo o espaço florestal estaria visível. A revisão da literatura referida no capítulo 2, permitiu observar os diversos modelos de segmentação existentes, podendo assim ter uma visão das opções que se poderiam tomar ao longo da dissertação. Também aludiu ao trabalho atualmente já desenvolvido nesta área no combate aos incêndios, que apesar de extenso os dados utilizados foram sempre bastante diferentes, como a deteção de zonas onde existira um incêndio já ativo, ou no caso da segmentação de imagens, imagens de satélite sobre o espaço florestal, algo que não seria o objeto de estudo desta investigação. Sendo assim, esta dissertação é muito dificilmente comparável aos casos de estudo demonstrados, pois em causa estão imagens de torres de vigilância, que podem ter algum ruído ou objeto que impossibilite a identificação de uma determinada zona urbana.

Os dados recebidos pelo INOV apresentavam imagens de seis espaços florestais e respetivas máscaras de segmentação elaboradas manualmente. Para a construção do protótipo de segmentação de imagens de espaços florestais foi necessário em primeiro lugar proceder a um processo de limpeza do ruído existente, como nevoeiro, algo que em certas imagens demonstrou ser insuficiente, continuando a não permitir ter uma visão clara das zonas urbanas e espaços florestais, dificultando a realização de qualquer modelo. Após o tratamento das imagens foram seguidos diversos processos de segmentação iniciais como segmentação por cor, luminância e contornos. No entanto, devido ao elevado nível de detalhe ou semelhanças de cor e até luminância idêntica entre zonas urbanas e espaços florestais estes apenas se revelaram eficazes em casos muito específicos. Com base na investigação realizada na revisão da literatura

os modelos restantes com mais hipótese de serem bem-sucedidos foram o modelo *K-Means* e a construção de uma rede neuronal convolucional que teria como base as máscaras já existentes, elaboradas manualmente. Com os dois modelos candidatos à automatização da segmentação de imagem selecionados, deu-se início à construção do protótipo do modelo *K-Means*, que incidu sobre as imagens de espaços florestais previamente tratadas quanto ao ruído existente.

Na construção deste modelo foi preferido a divisão por dois *clusters*, uma vez que após uma análise das imagens não haveria mais detalhe para além de espaços florestais, zonas urbanas e zona de céu, sendo que a zona de céu seria irrelevante na segmentação por esta já estar a ser considerada no algoritmo construído pelo INOV. Restariam apenas duas zonas a identificar, e a divisão por dois *clusters* permitiria que as zonas de indecisão de zona urbana ou espaço florestal pudessem ser resolvidas, não criando novas zonas segmentadas. Ao analisar os resultados obtidos verificou-se que este aparentou ser bastante efetivo em imagens onde a distinção entre zonas urbanas e espaços florestais era evidente, existindo, no entanto, a necessidade de mencionar qual seria o *cluster* a ser removido, algo que não foi possível automatizar para todos os casos, uma vez que em certas zonas a percentagem de zonas urbanas e espaços florestais era idêntica, e/ou as zonas urbanas estariam confundidas com zonas urbanas devido ao ruído existente. Isto levou a que o modelo em certos casos confundisse o *cluster* certo a eliminar, levando a uma segmentação errada, realizando uma segmentação exagerada, ou cobrindo as zonas florestas em vez das zonas urbanas. Posto isto partiu-se para a realização da rede neuronal convolucional onde estes problemas não existiram por esta se basear nas máscaras previamente desenhadas e não na análise da imagem em si, no entanto o número de dados seria insuficiente para uma rede neuronal convolucional fidedigna, optando-se pela divisão de cada imagem, em nove sub-imagens.

Nesta dissertação optou-se pela utilização da arquitetura de segmentação de imagem DeepLabV3 para a rede neuronal convolucional ResNet-101. Esta rede neuronal convolucional foi preferida pela profundidade que esta tem, sendo justificada esta escolha pelo nível de detalhe elevado existente nas imagens utilizadas. Sendo que se estava perante uma classificação binária foi utilizada a função *sigmoid* na arquitetura do modelo. Construída a arquitetura do modelo que iria iterar sobre as sub-imagens criadas existira então a necessidade da realização dos conjuntos de treino e validação, para isso foi necessário transformar as imagens pertencentes a estes dois conjuntos em *tensors*, de forma a estas poderem ser empregues no modelo construído anteriormente, preferiu-se ainda pela escolha do otimizador Adam.

Com a arquitetura do modelo construída e conjuntos de treino e validação criados, seria ainda necessário a utilização de métodos de avaliação deste, de forma a evidenciar a sua

qualidade. Como métodos de avaliação foram utilizados os métodos, Intersection over Union e Dice Coefficient, retratados na secção 4.2.4 Treino e Validação.

Na análise às máscaras, o modelo aparentou ser satisfatório, cobrindo parte das zonas urbanas desejadas, o mesmo não foi retratado nos métodos de avaliação elaborados, este resultado é motivado por diferentes causas, a principal causa existente poderia ser o baixo nível de informação nas sub-imagens elaboradas, com isso, foi procedido a duas novas segmentações desta vez com as imagens originais e outra com as imagens balanceadas. A primeira apesar de uma melhor avaliação, continuou a não demonstrar uma segmentação bem-sucedida. Causa desta fraca avaliação já não se encontra na informação da imagem, mas nas máscaras que foram elaboradas manualmente, que por vezes não marcam toda a zona urbana de forma detalhada, levando a que o treino realizado sobre as máscaras não seja o melhor possível. Já na segunda tentativa, com os dados balanceados, os resultados foram significativamente melhores, apresentando os valores médios 0,51 e 0,78 para os métodos de avaliação Intersection over Union e Dice Coefficient, respetivamente.

Relativamente ao objetivo da criação de um protótipo que visasse automatizar a segmentação de imagens de zonas urbanas em espaços florestais, este não se encontra totalmente automatizado, visto que existira a necessidade de intervenção humana para a retificação das zonas que ficariam por assinalar, sendo esta a preferência desejada, não correr o risco de marcar espaços florestais como zonas urbanas.

6.1. Limitações do trabalho

Parte das limitações nesta dissertação foram devidas ao nível de detalhe existente em certas imagens em comparação com outras, levando a que as abordagens fossem bastante abrangentes, assim como o ruído existente em parte das imagens. Esta última limitação dificultou a distinção das zonas desejadas, levando a que estas fossem confundidas entre si. Outra limitação terá sido o reduzido número de dados para a construção da rede neuronal convolucional, levando à necessidade da divisão da imagem original em 9, perdendo assim algum detalhe existente.

6.2. Trabalho futuro

Para a continuação deste trabalho foi considerado realizar uma seleção alargada dos dados, onde seriam removidas quaisquer imagens que fossem contraproducentes para os modelos elaborados. Após esta melhoria na seleção dos dados, podem ser criados novos modelos recorrendo a outras arquiteturas de segmentação de imagens, e a outras redes neuronais convolucionais, com vista a um melhor resultado. Seria ainda benéfico um tratamento das máscaras fornecidas pelo INOV, de forma a certificar que esta apenas cobriria zonas urbanas.

Referências

- [1] Divisão de Gestão do Programa de Fogos Rurais (2022) “8.º Relatório Provisório de Incêndios Rurais – 2022 – 1 de Janeiro a 15 de Outubro”, Instituto da Conservação da Natureza e das Florestas.
- [2] “CICLOPE - Sistema integrado de videovigilância de grandes áreas” [Online] Acedido em Março de 2022, em: <https://www.inov.pt/project/ciclope/index.html>.
- [3] N. Alzaza e A. Yaakub, (2011) “Student's Mobile Information Prototype for the Higher Education Environment”, American Journal of Economics and Business Administration.
- [4] “A Conceptual Explanation of Convolutional Neural Networks (CNN’s)” [Online] Acedido em Março de 2022, em: <https://towardsdatascience.com/a-conceptual-explanation-of-convolutional-neural-networks-cnns-ccd2e62f213b>.
- [5] S. Abdulateef e M. Salman, (2021) “A Comprehensive Review of Image Segmentation Techniques”, Iraqi Journal for Electrical and Electronic Engineering.
- [6] M. Jena, S. P. Mishra, e D. Mishra, (2018) “A survey on applications of machine learning techniques for medical image segmentation”, International Journal of Engineering & Technology.
- [7] S. Dubey e Y. Gupta, (2018) “Computational Comparison of Various Existing Edge Detection Techniques for Medical Images”, International Journal of Computing and applications.
- [8] S. Al-amri1, N.V. Kalyankar e Khamitkar S.D, (2010) “Image Segmentation by Using Threshold Techniques”, Journal of Computing, Vol. 2, ISSN 2151-9617.
- [9] M. D. Zeiler e R. Fergus (2014) “Visualizing and understanding convolutional networks”, ECCV.
- [10] L. Chen, G. Papandreou, K. Murphy e Alan L. Yuille (2017) “DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs”, arXiv:1606.00915v2.
- [11] S. Kumar (2020) [Online] Acedido em Novembro de 2022, em: <https://towardsdatascience.com/understanding-k-means-k-means-and-k-medoids-clustering-algorithms-ad9c9fbf47ca>.

- [12] S. Kumar (2021) [Online] Acedido em Novembro de 2022, em: <https://towardsdatascience.com/fuzzy-c-means-clustering-is-it-better-than-k-means-clustering-448a0aba1ee7>.
- [13] S. Kumar (2021) [Online] Acedido em Outubro de 2022, em: <https://prosancons.com/education/pros-and-cons-of-k-means-clustering>.
- [14] S. Kumar (2021) [Online] Acedido em Outubro de 2022, em: <https://researchhubs.com/post/ai/fundamentals/fuzzy-c-means.html>.
- [15] A. Janowczyk, S. Chandran, R. Singh, D. Sasaroli, G. Coukos, M. D. Feldman, e A. Madabhushi (2012), “HighThroughput Biomarker Segmentation on Ovarian Cancer Tissue Microarrays via Hierarchical Normalized Cuts,” IEEE transactions on biomedical engineering, Vol. 59, No. 5.
- [16] H. G. Kaganami e Z. Beij (2009), “Region Based Detection versus Edge Detection,” IEEE Transactions on Intelligent information hiding and multimedia signal processing, pp. 1217-1221.
- [17] P. Sidiropoulos, V. Mezaris, I. Kompatsiaris e J. Kittler (2012), “Differential Edit Distance: A Metric for Scene Segmentation Evaluation,” IEEE transactions on circuits and systems for video technology, Vol. 22, No. 6.
- [18] M. Kaur e P. Goyal (2013), “A Review on Region Based Segmentation”, International Journal of Science and Research (IJSR) 2319-7064.
- [19] G. Forlani, C. Nardinocchi, M. Scaioni e P. Zingaretti (2006), “Complete Classification of Raw LIDAR Data and 3D Reconstruction of Buildings”, Pattern Analysis and Applications.
- [20] R. Caubalejo, (2021) “Image Segmentation, Thresholding, Otsu’s and HSV Segmentation” [Online] Acedido em Junho de 2022, em: <https://towardsdatascience.com/image-segmentation-part-1-6e0ddb5f7f8a>.
- [21] “Image Thresholding” [Online] Acedido em Junho de 2022, em: <https://mathworks.com/discovery/image-thresholding.html>.

- [22] “OpenCv Image Thresholding” [Online] Acedido em Junho de 2022 em: https://docs.opencv.org/4.x/d7/d4d/tutorial_py_thresholding.html.
- [23] D.T. Anderson, J.C. Bezdek, M. Popescu, e J.M. Keller (2010), “Comparing Fuzzy, Probabilistic, and Possibilistic Partitions”, IEEE Transaction On Fuzzy Systems, Vol. 18, nº 5.
- [24] F. Shmmala e W. Ashour (2012), “Color Based Image Segmentation using Different Versions of K-Means in two Spaces”, Global Advanced Research Journal of Engineering, Technology and Innovation, Vol. 1, ISSN 2315-5124.
- [25] L. Chen, G. Papandreou, F. Schroff e H. Adam (2017), “Rethinking Atrous Convolution for Semantic Image Segmentation”, arXiv:1706.05587v3.
- [26] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, e B. Schiele (2016), “The cityscapes dataset for semantic urban scene understanding”, IEEE Conference on Computer Vision and Pattern Recognition.
- [27] X. Li, Z. Liu, P. Luo, C. C. Loy, e X. Tang (2017), “Not all pixels are equal: Difficulty-aware semantic segmentation via deep layer cascade”, arXiv:1704.01344.
- [28] C. Farabet, C. Couprie, L. Najman, e Y. LeCun, (2013), “Learning hierarchical features for scene labeling”, PAMI.
- [39] G. Lin, C. Shen e I. Reid (2015), “Efficient piecewise training of deep structured models for semantic segmentation”, arXiv:1504.01013.
- [30] T. Pohlen, A. Hermans, M. Mathias, e B. Leibe (2016), “Full resolution residual networks for semantic segmentation in street scenes”, arXiv:1611.08323.
- [31] G. Ghiasi e C. C. Fowlkes (2016), “Laplacian reconstruction and refinement for semantic segmentation”, arXiv:1605.02264.
- [32] G. Lin, A. Milan, C. Shen, e I. Reid (2016), “Refinenet: Multipath refinement networks with identity mappings for high resolution semantic segmentation”, arXiv:1611.06612.
- [33] X. Li, Z. Jie, W. Wang, C. Liu, J. Yang, X. Shen, Z. Lin, Q. Chen, S. Yan, e J. Feng (2017), “Foveanet: Perspective-aware urban scene parsing” arXiv:1708.02421.
- [34] I. Kreso, S. Segvic, e J. Krapac (2017), “Ladder-style dense nets for semantic segmentation of large natural images”, ICCV CVRSUAD workshop.

- [35] X. Jin, X. Li, H. Xiao, X. Shen, Z. Lin, J. Yang, Y. Chen, J. Dong, L. Liu, Z. Jie, J. Feng, e S. Yan (2017), “Video scene parsing with predictive feature learning”, ICCV, 2017.
- [36] R. Zhang, S. Tang, M. Lin, J. Li, e S. Yan (2017), “Global-residual and local-boundary refinement networks for rectifying scene parsing predictions”, IJCAI.
- [37] R. Zhang, S. Tang, Y. Zhang, J. Li, e S. Yan (2017), “Scale-adaptive convolutions for scene parsing”, ICCV.
- [38] F. Shen, R. Gan, S. Yan, e G. Zeng (2017), “Semantic segmentation via structured patch prediction, context crf and guidance crf”, CVPR.
- [39] P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou, e G. Cottrell (2017), “Understanding convolution for semantic segmentation”, arXiv:1702.08502.
- [40] R. Gadde, V. Jampani, e P. V. Gehler (2017), “Semantic video cnns through representation warping”, ICCV.
- [41] Z. Wu, C. Shen, e A. van den Hengel (2016), “Wider or deeper: Revisiting the resnet model for visual recognition”, arXiv:1611.10080.
- [42] H. Zhao, J. Shi, X. Qi, X. Wang, e J. Jia (2016), “Pyramid scene parsing network“, arXiv:1612.01105.
- [43] P. Raudaschl, (2015), "Evaluation of Segmentation Methods on Head and Neck", Auto-Segmentation Challenge 2015.
- [44] K. He, J. Sun e X. Tang (2011), “Single Image Haze Removal Using Dark Channel Prior”, IEEE Transactions on Pattern Analysis And Machine Intelligence, Vol. 33, N° 12.
- [45] B. Desai, U. Kushwaha e S. Jha (2020), “Image Filtering- Techniques, Algorithm and Applications”, Gis Science Journal, ISSN 1869-9391.