



UNIVERSIDADE  
LUSÓFONA

# Segmentação de incêndios florestais

## Trabalho Final de Curso

Francisco Alves, a22207107, Ciencia de Dados

Orientador: João Carvalho

Orientador: Houda Harkat

Departamento de Engenharia Informática e Sistemas de Informação

Universidade Lusófona, Centro Universitário de Lisboa

26/06/2025

[www.ulusofona.pt](http://www.ulusofona.pt)

## **Direitos de cópia**

Segmentação de incêndios florestais, Copyright de Francisco Perpétuo Alves, Universidade Lusófona.

A Escola de Comunicação, Arquitectura, Artes e Tecnologias da Informação (ECATI) e a Universidade Lusófona (UL) têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objectivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

# Resumo

Os incêndios florestais constituem um dos desastres naturais mais devastadores, causando impactos ambientais, sociais e económicos significativos. Estes eventos ameaçam ecossistemas inteiros, comprometem a biodiversidade, colocam vidas humanas e animais em risco e contribuem para o agravamento das alterações climáticas devido à emissão de gases com efeito de estufa. Assim, a deteção precoce e eficiente de incêndios é essencial para mitigar os seus danos e permitir intervenções rápidas e eficazes.

O desenvolvimento de incêndios florestais ocorre em duas fases principais: a emissão de fumo, que é o primeiro indicador visível do início de uma combustão, e, posteriormente, o aparecimento de chamas, que marcam um estado mais avançado e perigoso do fogo. A deteção de fumo é particularmente desafiante devido à sua natureza translúcida e à dispersão em condições ambientais adversas, como ventos ou luminosidade variável. Por outro lado, o fogo é mais fácil de identificar, mas quando visível, normalmente já representa uma situação crítica. Assim, um sistema de deteção eficaz deve ser capaz de identificar tanto o fumo quanto o fogo para facilitar uma intervenção atempada.

Esta trabalho propõe o desenvolvimento de um sistema avançado baseado em deep learning para a segmentação simultânea de fumo e fogo em imagens. Serão exploradas duas arquiteturas reconhecidas pela sua eficiência em tarefas de segmentação, a U-Net[1] e a ResUNet, ambas capazes de realizar classificações pixel-a-pixel em cenários complexos. Os modelos serão treinados e avaliados utilizando dois conjuntos de dados: o Corsican, um dataset público, e o Gestosa, um dataset privado, ambos contendo imagens captadas por drones e sensores térmicos.

O estudo também investigará a melhoria dessas arquiteturas por meio da otimização das funções de perda para aumentar a sua precisão e robustez, especialmente em condições ambientais desafiadoras. A otimização das arquiteturas visa melhorar a deteção de padrões complexos, como os contornos irregulares de fumo e fogo, e garantir uma maior eficácia em diferentes cenários.

Espera-se que os resultados desta trabalho demonstrem a viabilidade e a eficácia de sistemas baseados em deep learning para a deteção precoce de incêndios florestais. Esta abordagem representa um avanço significativo ao combinar a identificação de fumo e fogo numa única solução, com potencial para transformar as estratégias de monitorização, resposta a emergências e preservação ambiental.

**Palavras-chave:**Incêndios florestais; Deep learning, Segmentação de imagens; U-Net, ResUNet-a,ResUNet, Visão computacional.

# Abstract

Wildfires are one of the most devastating natural disasters, causing significant environmental, social and economic impacts. These events threaten entire ecosystems, compromise biodiversity, put human and animal lives at risk and contribute to worsening climate change due to greenhouse gas emissions. Therefore, early and efficient detection of fires is essential to mitigate their damage and enable quick and effective interventions.

Wildfires develop in two main phases: the emission of smoke, which is the first visible indicator of the start of combustion, and then the appearance of flames, which mark a more advanced and dangerous stage of the fire. Smoke detection is particularly challenging due to its translucent nature and dispersion in adverse environmental conditions, such as winds or variable lighting. On the other hand, fire is easier to identify, but when it is visible, it usually already represents a critical situation. Therefore, an effective detection system must be able to identify both smoke and fire to facilitate timely intervention.

This work proposes the development of an advanced system based on deep learning for the simultaneous segmentation of smoke and fire in images. Two architectures recognized for their efficiency in segmentation tasks will be explored, U-Net[1] and ResUNet, both capable of performing pixel-by-pixel classifications in complex scenarios. The models will be trained and evaluated using two datasets: Corsican, a public dataset, and Gestosa, a private dataset, both containing images captured by drones and thermal sensors.

The study will also investigate improving these architectures by optimizing the loss functions to increase their accuracy and robustness, especially in challenging environmental conditions. The optimization of the architectures aims to improve the detection of complex patterns, such as the irregular contours of smoke and fire, and ensure greater effectiveness in different scenarios.

The results of this work are expected to demonstrate the feasibility and effectiveness of deep learning-based systems for the early detection of Wildfires. This approach represents a significant advance by combining smoke and fire identification in a single solution, with the potential to transform monitoring, emergency response and environmental preservation strategies.

**Keywords:** Wildfires; Deep learning, Image segmentation; U-Net, ResUNet-a, ResUNet, Computer vision.

# Índice

|                          |           |
|--------------------------|-----------|
| <b>Resumo</b>            | <b>2</b>  |
| <b>Abstract</b>          | <b>3</b>  |
| <b>Índice</b>            | <b>4</b>  |
| <b>Lista de Figuras</b>  | <b>5</b>  |
| <b>Lista de Tabelas</b>  | <b>6</b>  |
| <b>1- Introdução</b>     | <b>7</b>  |
| <b>2- Estado de arte</b> | <b>10</b> |
| <b>3 - Metodologias</b>  | <b>18</b> |
| <b>4 - Resultados</b>    | <b>25</b> |
| <b>5 - Calendário</b>    | <b>28</b> |
| <b>6 - Conclusão</b>     | <b>30</b> |
| <b>Bibliografia</b>      | <b>32</b> |

# Lista de Figuras

|    |  |    |
|----|--|----|
| 1  | Perda de cobertura global de árvores devido a incêndios, 2001-2022 . . . .   | 10 |
| 2  | Exemplo de uma imagem do dataset . . . . .   | 11 |
| 3  | Máscara criada manualmente para a imagem do dataset . . . . .  | 11 |
| 4  | Arquitetura MaskSU R-CNN . . . . .   | 13 |
| 5  | Arquitetura geral de segmentação semântica de incêndios . . . . .  | 14 |
| 6  | Arquitetura U-Net . . . . .  | 15 |
| 7  | Arquitetura Unet . . . . .   | 19 |
| 8  | Arquitetura ResUNet-a . . . . .  | 21 |
| 9  | Análise visual da segmentação: (a) imagem original; (b) máscara de referência; (c) predição do modelo; (d) mapa de erro com falsos positivos (a amarelo) e falsos negativos (em azul). . . . . | 27 |
| 10 | Calendário de atividades do TFC . . . . .  | 28 |

## Lista de Tabelas

|   |   |    |
|---|---|----|
| 1 | Comparação de algoritmos para segmentação de incêndios florestais . . .                               | 16 |
| 2 | Resultados da U-Net com função de perda Cross Entropy e diferentes valores de learning rate . . . . . | 25 |
| 3 | Tabela de Resultados da Resunet para Diferentes Losses e Learning Rates                               | 26 |

# 1- Introdução

## 1.1 Motivação

Os incêndios florestais são uma ameaça cada vez maior em diversas regiões do mundo, causando impactos ambientais, prejudicando comunidades e afetando a economia local. Estes eventos devastadores têm sido amplificados pelas alterações climáticas, resultando em incêndios mais frequentes e intensos. O impacto ambiental é particularmente grave, com a destruição de vastas áreas de floresta, a perda de biodiversidade e a contribuição significativa para as emissões de gases com efeito de estufa, exacerbando o problema global das mudanças climáticas. Além disso, as comunidades afetadas por incêndios florestais enfrentam grandes desafios económicos e sociais, desde a destruição de bens e infraestruturas até à perda de vidas humanas.

A deteção precoce de incêndios florestais é um dos fatores mais críticos para minimizar estes danos. Identificar um foco de incêndio logo nas suas fases iniciais pode ser determinante para um combate eficaz e para a proteção de vidas e bens. Tradicionalmente, a deteção de incêndios tem sido realizada por meio de vigilância humana ou sistemas baseados em sensores. No entanto, estes métodos apresentam limitações em termos de cobertura e precisão, especialmente em áreas remotas ou de difícil acesso, onde a intervenção rápida é crucial.

Com o avanço das tecnologias de monitorização por satélite, drones e sensores, surgem novas oportunidades para melhorar a deteção de incêndios em tempo real. Uma dessas abordagens emergentes é o uso de técnicas de visão computacional combinadas com algoritmos de deep learning. Estes algoritmos, ao processarem grandes volumes de dados provenientes de imagens aéreas, podem identificar padrões visuais complexos, como o fumo e o fogo, de forma muito mais eficiente do que os métodos tradicionais. A segmentação de imagens, por exemplo, permite a separação precisa de áreas de interesse, como os focos de incêndio, facilitando a tomada de decisões rápidas e informadas.

A grande importância deste trabalho reside na exploração de soluções tecnológicas que possam aumentar a eficácia dos sistemas de deteção de incêndios florestais, contribuindo para a sua antecipação e combate. A utilização de imagens aéreas e algoritmos avançados oferece uma oportunidade única para superar as limitações dos métodos tradicionais, proporcionando uma abordagem mais eficaz e rápida para detetar focos de incêndio logo nas suas fases iniciais. A relevância deste trabalho é ainda maior considerando que a deteção precoce é fundamental para uma resposta eficaz e para a minimização dos danos ambientais e sociais causados por esses eventos.

## 1.2 Objetivo

O objetivo deste estudo é desenvolver e avaliar um sistema baseado em deep learning para a deteção de incêndios florestais em tempo real, com ênfase na redução do tempo de resposta e na melhoria da eficácia das operações de combate. A proposta visa usar modelos de aprendizagem profunda para identificar e localizar focos de fogo e fumo, informação crucial para uma resposta precoce e eficiente. Através da utilização de técnicas avançadas de segmentação de imagens, este trabalho procura aprimorar a capacidade de distinguir entre áreas queimadas, não afetadas e em risco, possibilitando uma análise detalhada e precisa.

Com isso, espera-se não apenas otimizar os processos de monitorização em tempo



real, mas também contribuir para a redução dos impactos ambientais, sociais e económicos provocados pelos incêndios florestais. A aplicação de soluções tecnológicas inovadoras permitirá um maior grau de automação e assertividade nas decisões de gestão de riscos, tornando o sistema mais eficaz na minimização dos danos causados por incêndios em áreas florestais e nas comunidades circundantes.

### **1.3 Pergunta de pesquisa**

É possível aumentar significativamente a precisão e a eficácia na deteção precoce de incêndios florestais utilizando técnicas de deep learning aplicadas a imagens de incêndios?

### **1.4 Hipótese**

A hipótese central deste estudo é que a aplicação de técnicas avançadas de deep learning, especificamente através de modelos de segmentação, pode melhorar substancialmente a deteção precoce e a precisão na identificação de incêndios florestais. A utilização de redes neuronais profundas, quando aplicadas a imagens aéreas ou de satélite, permite não só a segmentação eficaz dos focos de fogo, mas também a capacidade de detetar o fumo de maneira antecipada, o que é crucial para uma resposta rápida.

Espera-se que, ao otimizar os modelos existentes e adaptá-los a diferentes conjuntos de dados, seja possível alcançar uma solução robusta e precisa que identifique focos de incêndio e fumo com uma taxa mínima de falsos positivos e falsos negativos. A ideia é que, com a melhoria contínua das arquiteturas utilizadas, a deteção de incêndios se torne mais eficaz, rápida e adaptável a diferentes cenários, reduzindo significativamente o tempo de resposta e minimizando os danos causados pelos incêndios florestais.

Este estudo também investiga como a adaptação desses modelos pode levar a uma maior eficiência na utilização de recursos em sistemas automáticos de monitorização, além de contribuir para o desenvolvimento de tecnologias mais eficazes na mitigação dos impactos ambientais e sociais dos incêndios. A hipótese é que, ao aplicar e refinar esses modelos, será possível criar uma abordagem mais precisa e eficiente para a deteção e monitorização em tempo real de incêndios florestais.

### **1.5 Etapas do relatório**

O presente relatório está organizado em três capítulos principais, estruturados para apresentar, de forma clara e sistemática, os objetivos, os fundamentos teóricos e a abordagem prática adotada neste trabalho.

No Capítulo 1 - Introdução, são apresentados os aspetos motivadores deste estudo, evidenciando a importância e a relevância do tema abordado. Este capítulo também estabelece os objetivos gerais e específicos do trabalho, formula a pergunta de pesquisa e apresenta a hipótese proposta. Esses elementos fundamentam o problema em questão, oferecendo uma visão clara do propósito deste estudo e justificando sua realização no contexto científico e prático.

O Capítulo 2 - Estado da Arte, oferece uma revisão teórica e prática sobre os principais tópicos relacionados à pesquisa. Inicialmente, são descritos os principais datasets utilizados na área, com destaque para suas características, diversidade e representatividade. Em seguida, são exploradas as técnicas de anotação, detalhando os processos utilizados para criar um dataset de forma consistente e confiável. Posteriormente, são

analisados os algoritmos existentes, discutindo suas características, abordagens e desempenho. Por fim, é apresentada uma comparação entre os algoritmos, com base em métricas de avaliação e aplicações específicas, identificando as limitações e oportunidades para futuras melhorias.

O Capítulo 3 - Metodologia, descreve as abordagens utilizadas para implementar e avaliar os modelos propostos. Em particular, são exploradas as arquiteturas baseadas em U-Net e ResUNet-a, com explicações detalhadas das arquiteturas e como aumentar a eficácia da segmentação. São também discutidas técnicas para otimizar o treino dos modelos, como ajustes de funções de perda adequadas a problemas com classes desequilibradas.

Capítulo 4 - Resultados apresenta a análise dos experimentos realizados para avaliar o desempenho dos modelos de segmentação. São discutidos os impactos das diferentes funções de perda, taxas de aprendizagem e estratégias de regularização na qualidade da segmentação. Além disso, são exploradas as métricas utilizadas para a avaliação, permitindo uma comparação objetiva entre as abordagens testadas. A análise dos resultados busca identificar as melhores configurações para otimizar a segmentação, garantindo um modelo eficaz e generalizável.

O Capítulo 5 - Calendário, detalha as etapas organizacionais do trabalho, com base num cronograma estruturado que abrange desde a pesquisa científica até à redação do relatório. Este capítulo expõe o planeamento das atividades e discute as estratégias utilizadas para alcançar os objetivos de forma eficiente e no prazo estipulado. O calendário apresentado encontra-se atualizado até à data deste relatório, refletindo as etapas concluídas e os ajustes necessários para garantir a realização dos objetivos no período estabelecido.

O Capítulo 6 - Conclusão, apresenta as reflexões finais sobre o trabalho desenvolvido, destacando a relevância do tema e as contribuições do sistema proposto. Este capítulo discute os principais passos realizados durante o projeto, bem como as potencialidades da abordagem baseada em deep learning para a segmentação de fumo e fogo. Além disso, são abordados os benefícios esperados desta solução para a área de monitorização de incêndios florestais, concluindo com a importância do trabalho para avanços científicos e aplicações práticas.

## 2- Estado de arte

### O impacto e desafios dos incêndios florestais

Os incêndios florestais são uma ameaça crítica para os ecossistemas naturais, causando perdas significativas para a biodiversidade, degradação ambiental e impactando diretamente o bem-estar humano. As florestas, frequentemente chamadas de "pulmões do planeta", desempenham um papel essencial na filtragem do ar, absorvendo dióxido de carbono (CO<sub>2</sub>) e libertando oxigênio (O<sub>2</sub>), além de abrigar uma ampla diversidade de vida selvagem e proteger as culturas agrícolas contra eventos climáticos extremos (Zanchi et al., 2021 [2]; San-Miguel-Ayanz et al., 2013 [3]). A gravidade desses impactos pode ser observada na perda global de cobertura arbórea entre os anos de 2001 e 2022, apresentada na Figura 1. Os pontos vermelhos destacam as áreas afetadas por incêndios florestais, algumas das quais estão sob esforços de restauração significativos.

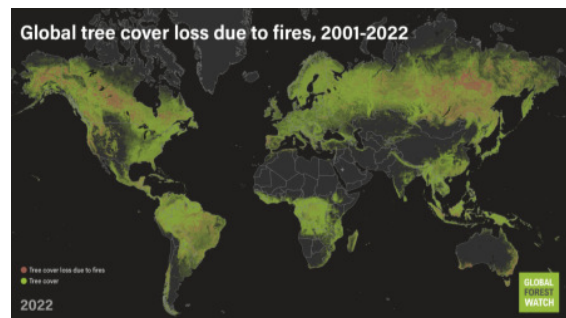


Figure 1: Perda de cobertura global de árvores devido a incêndios, 2001-2022

### 2.1 Datasets

O presente trabalho recorre a dois conjuntos de dados, um público e outro privado, utilizados na área de detecção de incêndios florestais: o Corsican[4] Dataset e o Firefront Gestosa Dataset. Estes datasets fornecem imagens de incêndios capturadas em diferentes condições, contribuindo significativamente para o treino e validação dos modelos de segmentação e detecção. Além disso, foi criado um dataset personalizado com imagens provenientes desses dois conjuntos de imagens, mas com a adição manual de máscaras de anotação detalhadas, que incluem a segmentação de áreas de fogo, fumo e fundo (ground truths). Estas máscaras serão feitas manualmente, assegurando um elevado nível de precisão, o que permitirá ao modelo aprender de forma mais eficaz e com maior adaptação às características específicas do cenário em análise.

#### 2.1.1 Datasets Description

##### Corsican Dataset

O Corsican Dataset[4] contém aproximadamente 2000 imagens de incêndios florestais capturadas sob diferentes configurações de câmaras, utilizando espectros visíveis e infravermelhos próximos. As imagens apresentam resolução de 1024 × 768 pixels e estão no formato PNG. Este conjunto de dados inclui imagens multimodais obtidas por câmaras

como a "JAI AD-080GE", capazes de captar simultaneamente espectros visíveis e infravermelhos através de óticas alinhadas. Cada imagem do dataset é acompanhada por uma máscara de segmentação correspondente, criada utilizando técnicas de homografia para registo preciso das imagens.

### **Firefront Gestosa Dataset**

O FireFront Gestosa Dataset é um conjunto de dados criado para apoiar pesquisas na detecção e segmentação de incêndios florestais. Ele contém imagens aéreas captadas em cenários reais, abrangendo uma variedade de condições ambientais e intensidades de fogo.

#### **2.1.2 Datasets annotation technique**

As máscaras de anotação, essenciais para o treino supervisionado dos modelos, foram criadas manualmente utilizando a ferramenta MATLAB ImageLabeler [5]. O processo de anotação envolveu a classificação de cada pixel das imagens como fogo, fumo ou fundo. A Figura 2 apresenta uma imagem do dataset de incêndios florestais, e a Figura 3 mostra a sua máscara feita manualmente.



Figure 2: Exemplo de uma imagem do dataset



Figure 3: Máscara criada manualmente para a imagem do dataset

## **2.2 Algoritmos Existentes**

A detecção de incêndios florestais utilizando imagens aéreas tem sido amplamente estudada com diferentes abordagens de algoritmos. Cada um desses algoritmos apresenta vantagens e desvantagens dependendo do tipo de imagem utilizada, das condições ambientais e dos objetivos do estudo. A seguir, apresentamos uma análise comparativa de

alguns estudos recentes que utilizam diferentes métodos para detetar incêndios florestais em imagens, com foco nas metodologias, resultados e limitações de cada um.

## **MaskSU R-CNN**

O artigo "Forest Fire Segmentation from Aerial Imagery Data Using an Improved Instance Segmentation Model"[6] apresenta o modelo MaskSU R-CNN, ilustrado na Figura 4, uma evolução do Mask R-CNN, desenvolvida especificamente para melhorar a segmentação de áreas de fogo em cenários florestais. A deteção e segmentação de regiões de fogo em imagens são tarefas cruciais para a prevenção e controlo de incêndios, mas enfrentam desafios significativos devido ao baixo contraste entre o fogo e o fundo, à natureza dinâmica das chamas e às condições ambientais adversas. O MaskSU R-CNN propõe uma solução que aborda estas limitações, integrando melhorias ao Mask R-CNN para alcançar uma segmentação mais precisa e eficiente.

Inspirado pela arquitetura U-Net [1], o MaskSU R-CNN incorpora uma ramificação MaskIoU em formato de "U", o que otimiza a segmentação de incêndios florestais ao preservar características importantes durante o processamento. O modelo beneficia de aprendizagem por transferência, utilizando pesos pré-treinados no dataset COCO. Este método acelera o processo de treino, aproveitando as características extraídas de um grande conjunto de dados diversificado para melhorar o desempenho do modelo em bases de dados específicas e menores, como as usadas na deteção de fogo. Durante o treino, pixels da imagem são selecionadas e classificadas como positivas ou negativas com base no limite de IoU (Intersection over Union), o que impacta diretamente a função de perda e direciona o modelo para uma segmentação mais robusta.

O treino do MaskSU R-CNN 4 é orientado por uma função de perda multibranch, composta por múltiplos componentes: perda RPN (Region Proposal Network), perda de classificação, perda de regressão de caixas delimitadoras, perda de máscara e perda de MaskIoU. Estes componentes trabalham de forma integrada para detetar e segmentar áreas de fogo com alta precisão. A perda RPN é responsável pela geração de propostas de regiões relevantes, enquanto a regressão de MaskIoU mede a qualidade da segmentação ao comparar as máscaras previstas com as máscaras de referência (ground truth). Este mecanismo é fundamental para otimizar a sobreposição entre as máscaras previstas e as reais, contribuindo para a melhoria contínua do modelo durante o treino.

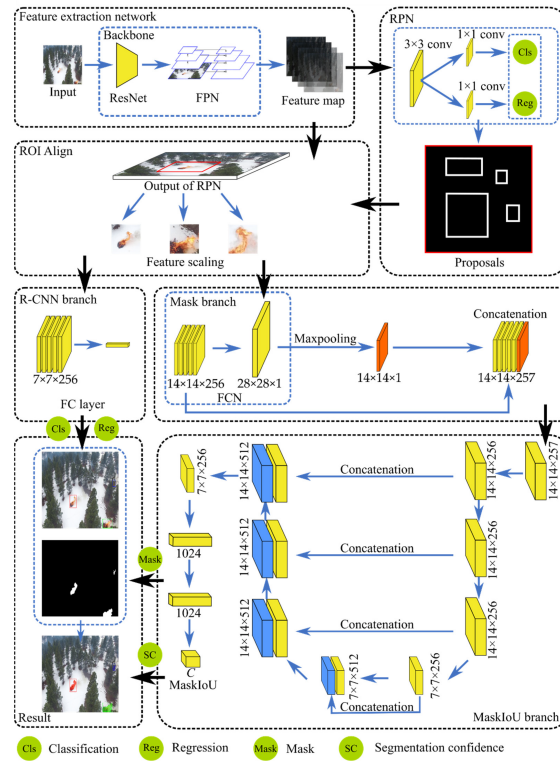


Figure 4: Arquitetura MaskSU R-CNN

A arquitetura equilibrada do MaskSU R-CNN permite ajustar os hiperparâmetros de forma a manter um equilíbrio eficaz entre classificação, regressão de caixas delimitadoras e segmentação de máscaras. Esta abordagem garante uma convergência estável durante o treino e resultados consistentes em diferentes cenários.

Os resultados apresentados no artigo demonstram a eficácia do MaskSU R-CNN, que alcança uma precisão de 91,85%, recall de 88,81%, F1-score de 90,30% e um mean intersection over union (mIoU) de 82,31%. Estes valores destacam a capacidade do modelo para realizar uma segmentação precisa e eficiente, mesmo em condições desafiadoras. Assim, o MaskSU R-CNN posiciona-se como uma ferramenta promissora para a detecção e monitorização de incêndios florestais, oferecendo suporte fundamental para a gestão e mitigação de desastres ambientais.

## GPG e MEI

O artigo "Optimized Deep Learning Model for Fire Semantic Segmentation"[7] apresenta uma abordagem para segmentação semântica de incêndios, ilustrada na figura 5, abordando desafios relacionados à identificação precisa de áreas afetadas pelo fogo. O trabalho propõe melhorias específicas em modelos de deep learning, com foco na correção de erros nas bordas das áreas em chamas e na captação de informações contextuais importantes para o cenário de incêndios.

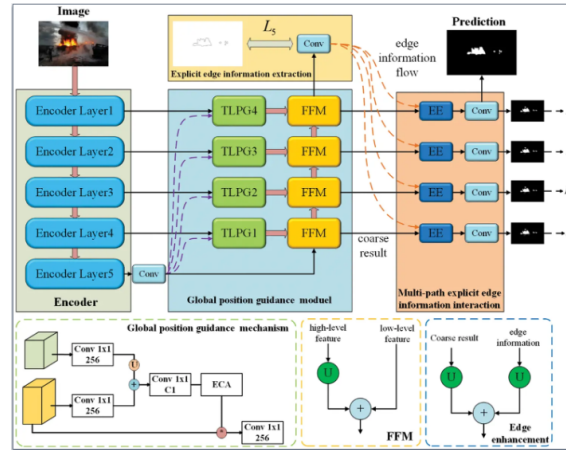


Figure 5: Arquitetura geral de segmentação semântica de incêndios

Os autores desenvolveram dois módulos principais para otimizar a segmentação. O Guia de Posição Global (Global Position Guidance - GPG) concentra-se na correção de falhas na identificação de características espaciais de baixo nível, oferecendo um mapeamento mais preciso da posição dos incêndios. O Módulo de Interação Explícita de Informações de Bordas (Multi-path Explicit Edge Information - MEI) tem como objetivo melhorar a definição e precisão das bordas das áreas de fogo, utilizando uma arquitetura multi-vias que integra informações de diferentes resoluções hierárquicas. Este design permite uma análise detalhada e consistente, mesmo em condições desafiadoras, como a presença de fumo ou reflexos.

Os modelos foram avaliados em bases de dados específicas de incêndios utilizando métricas amplamente reconhecidas no campo. Para a métrica Intersection over Union (IoU), o modelo alcançou um desempenho de 82,3%, superando métodos anteriores como DeepLabv3+ e U-Net. Além disso, obteve uma precisão global de 92,1%, demonstrando uma significativa melhoria na segmentação de áreas afetadas pelo fogo e na identificação de bordas com maior detalhe.

## U-net

O artigo "Aerial Imagery Pile Burn Detection Using Deep Learning: The FLAME Dataset"[8] apresenta uma metodologia para segmentação de imagens, ilustrada na figura 6, focada na identificação de áreas com fogo em imagens aéreas. A segmentação é tratada como um problema de classificação binária por pixel, visando gerar máscaras que identifiquem regiões ativamente em chamas. Essa abordagem é particularmente útil em cenários como a detecção de pequenos focos de fogo e o monitorização de áreas de queima.

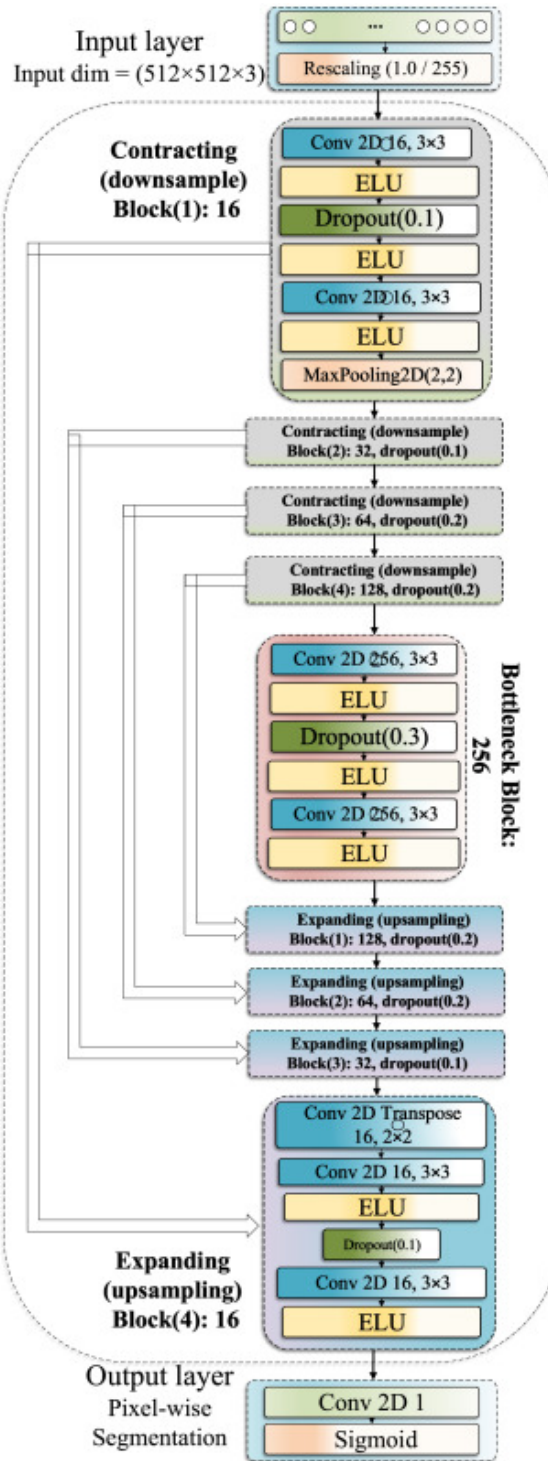


Figure 6: Arquitetura U-Net

Para alcançar esse objetivo, os autores utilizaram uma versão modificada da rede U-Net, conhecida por sua aplicação em segmentação de imagens biomédicas. No modelo adaptado, a função de ativação ReLU foi substituída pela Exponential Linear Unit (ELU), visando suavizar os resultados e melhorar a precisão. A arquitetura do modelo manteve a estrutura clássica da U-Net, composta por caminhos simétricos de contração e expansão conectados por operações de concatenação, o que preserva informações de alta resolução durante o processamento das imagens. Além disso, para minimizar problemas de overfitting devido à quantidade limitada de dados, foi aplicada a técnica de



dropout durante o treino.

O conjunto de dados utilizado foi o FLAME dataset, que contém imagens aéreas anotadas manualmente para a criação de máscaras que representam as regiões de fogo. A ferramenta MATLAB Image Labeler foi usada para gerar as anotações, e o dataset foi composto por 2003 frames. As imagens RGB foram normalizadas, com cada pixel classificado como "fogo" ou "não-fogo" (fundo). O modelo foi treinado com a função de perda binary cross entropy, otimizada pelo algoritmo Adam.

Os resultados obtidos demonstram a eficácia do modelo U-Net modificado. O desempenho foi avaliado com base em várias métricas: uma precisão de 91,99%, recall de 83,88%, área sob a curva (AUC) de 99,85%, F1-Score de 87,75%, sensibilidade de 83,12%, especificidade de 99,96% e Intersection over Union (IoU) de 78,17%. Esses resultados indicam uma capacidade robusta do modelo em segmentar regiões de fogo, oferecendo uma ferramenta valiosa para o monitorização e a gestão de incêndios em ambientes aéreos.

## 2.3 Comparação de Algoritmos Existentes

A Tabela 1 apresenta uma comparação entre diferentes algoritmos de detecção e segmentação de incêndios florestais, com base nos artigos analisados.

| Artigo   | Algoritmo(s) | Métricas (%)  |
|--|--------------|---|
| <i>Forest Fire Segmentation from Aerial Imagery Data Using an Improved Instance Segmentation Model [6]</i> | MaskSU R-CNN | <ul style="list-style-type: none"><li>• Precisão: 91,85%</li><li>• Recall: 88,81%</li><li>• F1-Score: 90,30%</li><li>• mIoU: 82,31%</li></ul>   |
| <i>Optimized Deep Learning Model for Fire Semantic Segmentation[7]</i>                                     | GPG e MEI    | <ul style="list-style-type: none"><li>• IoU: 82,3%</li><li>• Precisão global: 92,1%</li></ul>   |
| <i>Aerial Imagery Pile Burn Detection Using Deep Learning: The FLAME Dataset[8]</i>                        | U-Net        | <ul style="list-style-type: none"><li>• Precisão: 91,99%</li><li>• Recall: 83,88%</li><li>• AUC: 99,85%</li><li>• F1-Score: 87,75%</li><li>• Sensibilidade: 83,12%</li><li>• Especificidade: 99,96%</li><li>• IoU: 78,17%</li></ul> |

Table 1: Comparação de algoritmos para segmentação de incêndios florestais

## 2.4 Benchmarking

A aplicação de tecnologias avançadas de deep learning para a detecção de incêndios florestais tem vindo a crescer em todo o mundo, incluindo na Europa, onde iniciativas relevantes mostram como estas abordagens podem ser adaptadas a diferentes contextos ambientais. Este subcapítulo explora implementações relevantes de sistemas baseados em inteligência artificial em países europeus.

### Sistema EFFIS (União Europeia)

O *European Forest Fire Information System (EFFIS)*[9] é uma iniciativa da Comissão Europeia, criada para monitorizar e prever incêndios florestais em toda a Europa. O sistema utiliza principalmente dados de satélites para fornecer informações em tempo real sobre incêndios florestais e os riscos associados. Além disso, o EFFIS funciona como uma plataforma colaborativa que permite a troca de dados entre países da União Europeia, melhorando a capacidade de resposta coletiva no combate a incêndios.

### Sistema CICLOPE (Portugal)

O *CICLOPE*[10] é um sistema de monitorização de incêndios florestais desenvolvido em Portugal, com foco na detecção precoce de focos de incêndio em áreas de risco elevado, como florestas e regiões de difícil acesso. Este sistema utiliza imagens de satélite e tecnologias de inteligência artificial para detetar incêndios em estágio inicial, permitindo a identificação rápida de focos emergentes e a emissão de alertas em tempo real. A tecnologia do CICLOPE tem sido utilizada em várias regiões de Portugal para melhorar a resposta a incêndios e reduzir os danos ambientais. A aplicação de inteligência artificial permite uma análise precisa das imagens e dados, contribuindo para uma gestão mais eficiente e rápida dos incêndios florestais.

### Relevância destas Implementações para Portugal

As soluções implementadas em países da União Europeia, como o EFFIS e o CICLOPE, demonstram como tecnologias avançadas de monitorização e análise preditiva podem melhorar a gestão de incêndios florestais. A aplicação dessas tecnologias em Portugal já tem contribuído significativamente para a detecção precoce e a resposta rápida a incêndios, fortalecendo as estratégias de prevenção e mitigação. Esses sistemas são fundamentais para otimizar a gestão de recursos, melhorar a coordenação entre as autoridades e reduzir os danos ambientais e sociais causados pelos incêndios. pactos ambientais e humanos causados pelos incêndios florestais.

## 3 - Metodologias

### 3.1 Metodologia

A segmentação de focos de incêndio em imagens representa um desafio significativo devido à complexidade dos padrões visuais envolvidos. Elementos como a interferência do fumo, a translucidez das chamas e a variabilidade dos cenários ambientais dificultam a identificação precisa dessas regiões. Métodos tradicionais de processamento de imagens, baseados em modelagem matemática ou redes neurais superficiais, frequentemente apresentam limitações na generalização e precisão, resultando em um elevado número de falsos positivos e falsos negativos. Em contraste, as redes neurais profundas, especialmente as Redes Neurais Convolucionais (CNNs), destacam-se pela sua capacidade de aprender representações robustas e características discriminativas diretamente dos dados, superando essas limitações.

Neste estudo, foram exploradas as arquiteturas U-Net e ResUNet-a para a segmentação de fogo e fumo. Essas redes são amplamente reconhecidas pela sua eficiência em preservar informações críticas durante o processamento de imagens e por sua capacidade de capturar detalhes em múltiplas escalas. A U-Net, composta por caminhos simétricos de contração e expansão conectados por operações de concatenação, possibilita a extração e a fusão eficaz de características em diferentes resoluções. A ResUNet-a, por sua vez, introduz blocos residuais, que aprimoram o fluxo de gradientes e mitigam problemas como o *vanishing gradient*, tornando o treinamento mais estável e eficaz na segmentação de padrões visuais complexos.

Além das vantagens estruturais da ResUNet-a, sua robustez será avaliada por meio de experimentos com diferentes configurações de treinamento. Para isso, serão analisadas quatro funções de perda — CrossEntropy, Dice Loss, Tversky Loss, Jaccard Loss, Generalized Dice e Hausdorff loss — combinadas com três taxas de aprendizado distintas ( $10^{-4}$ ,  $10^{-3}$  e  $10^{-2}$ ). Essa abordagem permitirá examinar como cada combinação de *loss function* e *learning rate* influencia o desempenho da rede, garantindo uma segmentação eficiente dos focos de incêndio.

Após a definição dos melhores hiperparâmetros de *loss* e *learning rate*, a arquitetura ResUNet-A será aprimorada com a introdução de mecanismos de regularização para reduzir o overfitting e melhorar a capacidade de generalização do modelo.

O pipeline de implementação segue as etapas fundamentais de treino e validação dos modelos, utilizando dados anotados manualmente para a criação de máscaras que representam as áreas de interesse. O desempenho das arquiteturas U-Net e ResUNet-a será avaliado com base em métricas quantitativas como *Global Accuracy*, *Balanced Accuracy*, *F1-score* e *Intersection over Union* (IoU), permitindo uma comparação detalhada da eficácia dos modelos. A análise dos resultados possibilitará a identificação da configuração mais eficiente para a segmentação automática de incêndios, contribuindo para o avanço das técnicas de monitorização e detecção precoce de focos de fogo.

#### U-Net

O **U-Net** [1] é uma arquitetura de rede neuronal convolucional concebida para segmentação de imagens. O seu design em forma de "U" combina um **caminho de contração** (encoder), que extrai características de alto nível reduzindo progressivamente a resolução da imagem, e um **caminho de expansão** (decoder), que recupera a resolução original, preservando detalhes espaciais essenciais para a segmentação precisa.

Na implementação desenvolvida 7, a arquitetura segue essa estrutura fundamental, mas com ajustes específicos para a segmentação de fumo e fogo. O modelo é composto pelos seguintes componentes principais:

- **Blocos de Convolução Dupla (DoubleConv):** Cada camada do encoder e do decoder contém duas convoluções seguidas de ativações ReLU, melhorando a capacidade de extração de características.
- **Encoder:** Composto por quatro níveis de convolução dupla, intercalados com operações de max pooling, reduzindo a dimensionalidade da imagem enquanto aumenta a profundidade da representação.
- **Bottleneck:** Representa o ponto de menor resolução da rede, onde a informação extraída atinge a máxima abstração antes da reconstrução.
- **Decoder:** Inclui camadas de convolução transposta para upsampling da imagem, seguidas de convoluções duplas que refinam as características recuperadas.
- **Skip Connections:** Ligam as camadas correspondentes do encoder e do decoder, garantindo a preservação de detalhes importantes para a segmentação.
- **Camada Final:** Uma convolução  $1 \times 1$  para gerar o mapa segmentado, com tantas saídas quanto classes a serem segmentadas.

Essas características tornam o modelo altamente eficaz para a segmentação pixel a pixel, preservando os contornos das regiões segmentadas – um fator essencial na identificação de chamas e fumaça. Além disso, a rede pode ser otimizada por funções de perda como Cross-Entropy ou Dice Loss, enquanto o otimizador Adam é utilizado para acelerar a convergência do processo de otimização do modelo.

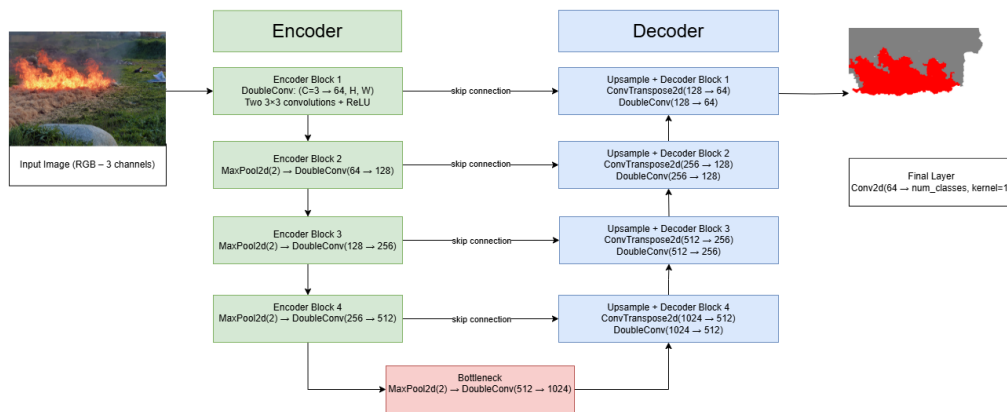


Figure 7: Arquitetura U-Net

## ResUNet-A

A **arquitetura ResUNet-A** [11] é uma evolução da arquitetura ResUNet, desenvolvida para tarefas complexas de segmentação, como a detecção de incêndios florestais em imagens de alta resolução. Esta arquitetura combina os **blocos residuais** da ResNet com a estrutura clássica do **U-Net**, introduzindo componentes adicionais para melhorar a extração de características e a preservação da informação espacial.

A arquitetura do **ResUNet-A** 8 é composta pelos seguintes componentes principais:

- **Encoder:** Tem como objetivo extrair características de baixo para alto nível da imagem de entrada. A imagem começa com 3 canais (RGB) e as dimensões espaciais são progressivamente reduzidas à medida que a imagem passa pelas camadas convolucionais. Isso permite capturar padrões em diferentes escalas. Cada camada do encoder é composta por blocos residuais, que garantem maior estabilidade no treinamento e facilitam a propagação dos gradientes em redes profundas. Nesta implementação, os blocos residuais do encoder utilizam diferentes valores de *stride* e *dilatação*, aumentando o campo receptivo da rede sem uma perda excessiva de resolução espacial.
- **Blocos Residuais:** Os blocos residuais são a base da arquitetura ResUNet-A. Cada bloco é composto por duas camadas convolucionais seguidas de normalização em lotes (Batch Normalization) e uma função de ativação ReLU. O principal objetivo dos blocos residuais é permitir que a rede aprenda funções de identidade, facilitando a propagação de gradientes e minimizando o problema de degradação que ocorre em redes profundas.

No código, a implementação do bloco residual é feita através da classe `ResidualBlock`. A estrutura do bloco inclui:

- **Primeira Convolução:** A primeira camada convolucional aplica um filtro de tamanho  $3 \times 3$ , com *stride* e *dilatação* ajustáveis (dependendo da profundidade do bloco). A operação é seguida de uma camada de normalização em lotes (Batch Normalization) e uma ativação ReLU.
- **Segunda Convolução:** A segunda camada convolucional aplica outro filtro de  $3 \times 3$ , com *stride* 1. Também é seguida por Batch Normalization.
- **Skip Connection:** A principal característica dos blocos residuais é a presença da *skip connection*, ou conexão de atalho. Essa conexão permite que a entrada original seja somada diretamente ao resultado da segunda convolução. Se o número de canais ou o *stride* for diferente, uma convolução  $1 \times 1$  é usada para ajustar as dimensões.
- **Ativação ReLU Final:** Após a soma da entrada com a saída das convoluções, aplica-se uma função de ativação ReLU para garantir que o modelo aprenda representações não lineares. O uso da ativação ReLU ajuda a mitigar o problema do gradiente explosivo ou de desaparecimento durante o treinamento.

A função `forward` na classe `ResidualBlock` descreve como o dado flui através de cada uma dessas operações e, ao final, realiza a soma da entrada com a saída do bloco convolucional. Essa soma é essencial para a efetividade do bloco residual, permitindo que a rede aprenda representações mais robustas, preservando as informações relevantes durante o processo de treinamento.

- **Bottleneck:** É a parte mais profunda da rede, onde a resolução da imagem é reduzida ao seu mínimo e a quantidade de informação extraída é maior. Durante esta fase, o número de canais atinge o valor máximo (1024), permitindo capturar as características mais complexas e abstratas da imagem. O bottleneck também é composto por um bloco residual, garantindo a consistência das representações ao longo da rede.
- **Decoder:** O principal objetivo do decoder é restaurar as dimensões espaciais da imagem até a sua resolução original, recuperando as informações perdidas durante o downsampling. O decoder utiliza camadas de **convoluções transpostas**

(ConvTranspose2d) para realizar o *upsampling*. Após cada operação de upsampling, é aplicada uma função `F.interpolate` para garantir que a dimensão espacial coincida com a das saídas do encoder (necessário para realizar a soma). Em seguida, realiza-se uma **soma** (e não concatenação como na U-Net clássica) com os mapas de características correspondentes do encoder. Cada bloco de upsampling é seguido de um bloco residual, que ajuda a refinar as características reconstruídas.

- **Camada Final:** A camada final do modelo é uma convolução  $1 \times 1$ , que mapeia a saída do último bloco residual para o número de classes da tarefa de segmentação. No caso da segmentação de incêndios florestais, essa camada gera uma máscara com três classes: 0 para o *background*, 1 para o *fumo* e 2 para o *fogo*. Cada pixel da imagem é classificado de acordo com uma dessas três categorias, permitindo a identificação das áreas afetadas por fogo, fumo ou as regiões de fundo.

O fluxograma 8 abaixo ilustra a sequência de operações na arquitetura ResUNet-A. Este mostra o fluxo de dados desde a entrada da imagem até a produção da saída segmentada, incluindo os processos de downsampling no encoder, o bottleneck e o upsampling no decoder, com as respectivas skip connections entre as camadas.

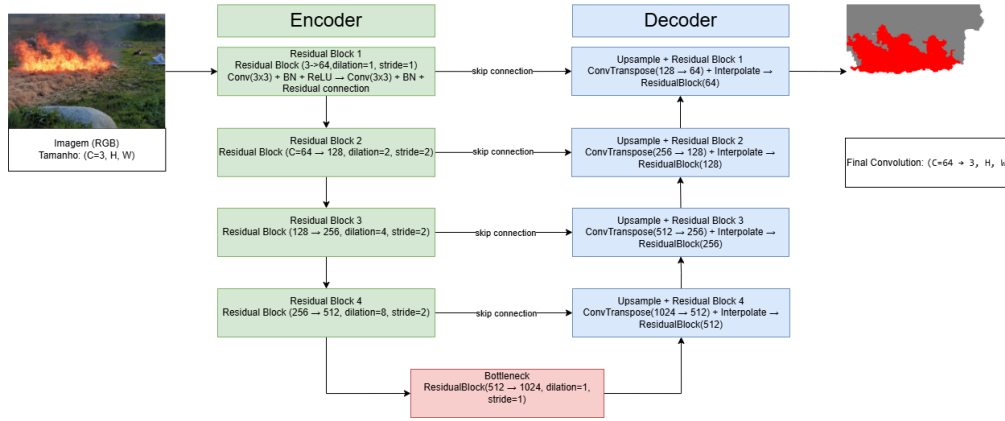


Figure 8: Arquitetura ResUNet-a

## 3.2 Métricas de avaliação

As métricas utilizadas para avaliar o desempenho dos modelos de segmentação incluem: *Global Accuracy*, *Balanced Accuracy*, *IoU* (Jaccard Score) e *F1 Score*.

- **Global Accuracy:** A Global Accuracy [12] fornece uma visão geral sobre a percentagem de pixels corretamente classificados, independentemente da classe a que pertencem. Ela é calculada como a razão entre o número de pixels corretamente classificados e o total de pixels na imagem. Embora amplamente utilizada pela sua simplicidade, a *Global Accuracy* pode ser influenciada por um desbalanceamento nas classes, o que pode tornar a métrica menos confiável em cenários com grande desequilíbrio entre as classes. A fórmula para o cálculo da *Global Accuracy* é dada por:

$$\text{Global Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

onde  $TP$  são os verdadeiros positivos,  $TN$  são os verdadeiros negativos,  $FP$  são os falsos positivos e  $FN$  são os falsos negativos. *Global Accuracy* oferece uma visão geral do desempenho, mas não distingue as classes de forma individual.

- **Balanced Accuracy:** A *Balanced Accuracy* [13] resolve o problema da *Global Accuracy* em cenários com classes desbalanceadas. Ela é calculada como a média da *accuracy* de cada classe, proporcionando uma medida mais justa do desempenho do modelo, especialmente quando a distribuição das classes na base de dados é desigual. A *Balanced Accuracy* é definida como:

$$\text{Balanced Accuracy} = \frac{1}{C} \sum_{i=1}^C \frac{TP_i}{TP_i + FN_i} \quad (2)$$

onde  $C$  é o número total de classes,  $TP_i$  é o número de verdadeiros positivos da classe  $i$  e  $FN_i$  é o número de falsos negativos da classe  $i$ .

- **IoU (Intersection over Union) - Jaccard Score:** A métrica *IoU* [14], também conhecida como *Jaccard Index*, mede a sobreposição entre as áreas previstas e as áreas reais. Ela é definida como a razão entre a interseção e a união das áreas previstas e reais. A *IoU* é calculada como:

$$\text{IoU} = \frac{TP}{TP + FP + FN} \quad (3)$$

onde  $TP$  representa os verdadeiros positivos,  $FP$  os falsos positivos e  $FN$  os falsos negativos. Esta métrica penaliza mais fortemente os falsos positivos e negativos, sendo útil para medir a precisão da segmentação.

- **F1 Score:** O *F1 Score* [15] é a média harmônica entre a precisão e o recall, fornecendo uma única métrica que combina tanto a precisão quanto a sensibilidade do modelo. Ele é especialmente útil em cenários onde é importante equilibrar a precisão e a capacidade do modelo de identificar todas as instâncias de uma classe. O *F1 Score* é definido como:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

onde  $\text{Precision} = \frac{TP}{TP+FP}$  e  $\text{Recall} = \frac{TP}{TP+FN}$ . O *F1 Score* assume um valor entre 0 e 1, sendo 1 o melhor desempenho possível.

### 3.3 Loss Functions

No contexto de treino de redes neurais profundas para tarefas de segmentação semântica, a escolha da função de perda (*loss function*) é um fator determinante para o desempenho do modelo. A função de perda mede a discrepância entre as previsões da rede e os valores reais (rótulos) do conjunto de dados, permitindo o ajuste iterativo dos pesos da rede através da propagação do erro e da otimização por redução do gradiente.

A segmentação de imagens apresenta desafios específicos devido à elevada desproporção entre classes (por exemplo, poucas regiões de fogo comparadas com grandes áreas sem fogo) e à necessidade de uma delimitação precisa das fronteiras dos objetos segmentados. Para mitigar esses desafios, utilizam-se diferentes funções de perda, cada uma com propriedades distintas que afetam a convergência e a precisão do modelo.

Neste estudo, foram avaliadas quatro funções de perda distintas: *Cross-Entropy Loss*, *Dice Loss*, *Teversky Loss* e *Jaccard Loss*. Estas funções foram escolhidas devido à sua capacidade de lidar com desbalanceamento de classes e à sua adequação para tarefas de segmentação.

- **Cross-Entropy Loss**

A *Cross-Entropy Loss* [16] é uma das funções de perda mais utilizadas para tarefas de classificação e segmentação multiclasse. Neste caso, onde o objetivo é distinguir entre fundo, fogo e fumo, a cross-entropy mede a discrepância entre as probabilidades previstas pela rede e os rótulos reais, atribuindo uma penalização maior a previsões com baixa confiança. Apesar da sua eficácia geral, esta função de perda não considera diretamente o desbalanceamento de classes, tratando todas com igual peso. Assim, pode ser menos eficaz na deteção de regiões pequenas, como o fogo e o fumo, que tendem a ser sub-representadas nos dados.

A *Cross-Entropy Loss* é definida como:

$$\mathcal{L}_{CE} = - \sum_i y_i \log(\hat{y}_i) \quad (5)$$

onde  $y_i$  representa a classe verdadeira para o pixel  $i$ , e  $\hat{y}_i$  a probabilidade prevista pelo modelo para essa classe.

- **Dice Loss**

A *Dice Loss* [17] baseia-se no coeficiente de Dice, uma métrica comum em segmentação de imagens, sendo particularmente útil quando há um forte desbalanceamento entre classes. Por exemplo, em cenários que a deteção de incêndios, onde as áreas de interesse (fogo e fumo) são significativamente menores que o fundo. A *Dice Loss* maximiza a sobreposição entre a máscara predita e a verdade de solo, favorecendo uma segmentação mais precisa das regiões minoritárias.

A *Dice Loss* é definida como:

$$\mathcal{L}_{Dice} = 1 - \frac{2 \sum y_i \hat{y}_i}{\sum y_i + \sum \hat{y}_i} \quad (6)$$

onde  $\sum y_i \hat{y}_i$  representa a interseção entre as previsões e os rótulos verdadeiros, enquanto  $\sum y_i + \sum \hat{y}_i$  representa a soma total das áreas segmentadas.

- **Tversky Loss**

A *Tversky Loss* [17] expande a *Dice Loss* ao introduzir parâmetros que permitem um controlo mais refinado sobre a penalização dos falsos positivos e falsos negativos. Isto é particularmente relevante para a segmentação de incêndios, onde erros podem ter impactos significativos na interpretação das imagens. A flexibilidade desta função de perda permite otimizar o modelo para um melhor compromisso entre sensibilidade e precisão, melhorando a segmentação das regiões críticas.

A *Tversky Loss* é definida como:

$$\mathcal{L}_{Tversky} = 1 - \frac{\sum y_i \hat{y}_i}{\sum y_i \hat{y}_i + \alpha \sum y_i (1 - \hat{y}_i) + \beta \sum (1 - y_i) \hat{y}_i} \quad (7)$$

onde  $\alpha$  controla a penalização sobre os falsos negativos e  $\beta$  sobre os falsos positivos. Valores mais elevados de  $\alpha$  favorecem a sensibilidade da segmentação, enquanto valores mais elevados de  $\beta$  favorecem a precisão.

- **Jaccard Loss**

A *Jaccard Loss* [17], também conhecida como *Intersection over Union (IoU) Loss*, mede a proporção entre a interseção e a união entre a predição e a verdade de solo. Esta função de perda é amplamente utilizada em tarefas de segmentação porque



penaliza mais fortemente predições que apresentam sobreposição reduzida com as regiões reais. No contexto da segmentação de incêndios, a *Jaccard Loss* é particularmente útil para garantir que o modelo não apenas deteta áreas de fogo e fumo, mas que também define corretamente os seus contornos, reduzindo ambiguidades na segmentação.

A *Jaccard Loss* é definida como:

$$\mathcal{L}_{Jaccard} = 1 - \frac{\sum y_i \hat{y}_i}{\sum y_i + \sum \hat{y}_i - \sum y_i \hat{y}_i} \quad (8)$$

onde o denominador representa a união entre as áreas previstas e as áreas reais, garantindo que a penalização por predições erradas seja proporcional ao erro cometido.

- **Generalized Dice Loss**

A *Generalized Dice Loss* [18] é uma extensão da *Dice Loss* que atribui pesos às diferentes classes com base na sua frequência, tornando-a mais robusta para cenários com desbalanceamento extremo. Em tarefas de segmentação de incêndios, onde o fogo e o fumo ocupam áreas significativamente menores do que o fundo, esta função de perda reduz a influência do desbalanceamento na otimização do modelo.

A *Generalized Dice Loss* é definida como:

$$\mathcal{L}_{GDL} = 1 - 2 \frac{\sum w_c \sum y_i^c \hat{y}_i^c}{\sum w_c \sum (y_i^c + \hat{y}_i^c)} \quad (9)$$

onde  $w_c$  é um peso atribuído a cada classe, normalmente definido como  $w_c = \frac{1}{(\sum y_i^c)^2}$ , equilibrando a contribuição de cada classe na perda total.

## 4 - Resultados

A análise dos resultados obtidos na segmentação de fogo e fumo foi conduzida a partir das métricas *Global Accuracy*, *Balanced Accuracy*, *F1-score* e *Intersection over Union* (IoU). Os testes foram realizados para três valores distintos de *learning rate* ( $10^{-4}$ ,  $10^{-3}$  e  $10^{-2}$ ), com o objetivo de identificar a melhor configuração para a tarefa de segmentação.

Inicialmente, avaliou-se o desempenho da arquitetura U-Net utilizando apenas a função de perda Cross Entropy, de forma a estabelecer uma linha de base para comparação com a arquitetura da Reunet. Os resultados desta fase encontram-se na Tabela 2. Posteriormente, foi realizada uma análise mais aprofundada com a arquitetura Resunet, testando múltiplas funções de perda (Cross Entropy, Dice Loss, Tversky Loss, Jaccard Loss e Generalized Dice), conforme apresentado na Tabela 3. Esta abordagem permitiu selecionar o modelo mais promissor e, em seguida, explorar o impacto das diferentes funções de perda e learning rates no seu desempenho.

| Learning Rate | Métrica           | Cross Entropy |
|---------------|-------------------|---------------|
| $10^{-4}$     | Accuracy          | 0.9871        |
|               | Balanced Accuracy | 0.9570        |
|               | F1 Score          | 0.9571        |
|               | IoU               | 0.9179        |
| $10^{-3}$     | Accuracy          | 0.9485        |
|               | Balanced Accuracy | 0.9495        |
|               | F1 Score          | 0.9486        |
|               | IoU               | 0.9025        |
| $10^{-2}$     | Accuracy          | 0.7100        |
|               | Balanced Accuracy | 0.7145        |
|               | F1 Score          | 0.7101        |
|               | IoU               | 0.5538        |

Table 2: Resultados da U-Net com função de perda Cross Entropy e diferentes valores de learning rate

| Learning Rate        | Métricas          | Cross Entropy | Dice Loss | Tversky Loss | Jaccard Loss | GeneralizedDice |
|----------------------|-------------------|---------------|-----------|--------------|--------------|-----------------|
| $1.0 \times 10^{-4}$ | Accuracy          | 0.9651        | 0.9477    | 0.9492       | 0.9638       | 0.9144          |
|                      | Balanced Accuracy | 0.9676        | 0.9490    | 0.9512       | 0.9661       | 0.9197          |
|                      | F1 Score          | 0.9652        | 0.9478    | 0.9492       | 0.9638       | 0.9144          |
|                      | IoU               | 0.9327        | 0.9009    | 0.9035       | 0.9302       | 0.8426          |
| $1.0 \times 10^{-3}$ | Accuracy          | 0.9649        | 0.9459    | 0.9455       | 0.9632       | 0.9093          |
|                      | Balanced Accuracy | 0.9674        | 0.9466    | 0.9482       | 0.9652       | 0.9158          |
|                      | F1 Score          | 0.9650        | 0.9460    | 0.9456       | 0.9632       | 0.9094          |
|                      | IoU               | 0.9323        | 0.8975    | 0.8969       | 0.9290       | 0.8342          |
| $1.0 \times 10^{-2}$ | Accuracy          | 0.9601        | 0.9290    | 0.7107       | 0.9515       | 0.8520          |
|                      | Balanced Accuracy | 0.9616        | 0.9314    | 0.7546       | 0.9535       | 0.8581          |
|                      | F1 Score          | 0.9601        | 0.9289    | 0.7152       | 0.9515       | 0.8506          |
|                      | IoU               | 0.9234        | 0.8674    | 0.5577       | 0.9081       | 0.7415          |

Table 3: Tabela de Resultados da Resunet para Diferentes Losses e Learning Rates

#### 4.1 Impacto da Função de Perda

A comparação entre as funções de perda revelou diferenças quantitativas relevantes no desempenho do modelo, embora todas tenham alcançado valores globais considerados elevados. A *Cross Entropy* destacou-se pela consistência dos seus resultados, obtendo as melhores ou segundas melhores métricas em praticamente todos os cenários, com destaque para a IoU de 0.9327 com *learning rate* de  $10^{-4}$ .

A *Jaccard Loss* também apresentou desempenho competitivo, com métricas muito próximas das da *Cross Entropy*, especialmente em *IoU* e *F1-score*. A *Tversky Loss* apresentou um comportamento mais instável: enquanto com  $10^{-4}$  alcançou uma *IoU* elevada de 0.9035 (superior à da *Dice Loss* e *Generalized Dice*), com  $10^{-2}$  essa mesma métrica desceu drasticamente para 0.5577, refletindo sensibilidade acentuada à variação da taxa de aprendizagem.

As funções *Dice Loss* e *Generalized Dice* apresentaram, de forma geral, desempenhos ligeiramente inferiores. A *Generalized Dice*, em particular, foi consistentemente a função com os valores mais baixos de *IoU* em todos os *learning rates*. Por exemplo, para  $10^{-4}$ , obteve *IoU* de 0.8426, consideravelmente inferior ao valor obtido com *Cross Entropy* no mesmo cenário.

Portanto, observa-se que, embora nenhuma função de perda tenha comprometido de forma significativa a capacidade do modelo, o desempenho varia conforme a taxa de aprendizagem utilizada. Isso indica que a escolha da função de perda deve ser feita em conjunto com a calibragem adequada dos hiperparâmetros, em vez de forma isolada.

#### 4.2 Influência da Taxa de Aprendizagem

A taxa de aprendizagem teve um papel importante na estabilidade e qualidade do treinamento. As menores taxas testadas ( $10^{-4}$  e  $10^{-3}$ ) proporcionaram resultados mais estáveis e elevados em todas as métricas. As variações entre essas duas taxas foram geralmente pequenas, o que sugere que o modelo apresenta robustez dentro dessa faixa.

No entanto, ao utilizar a taxa de  $10^{-2}$ , houve degradação significativa do desempenho em várias funções de perda, em especial na *Tversky Loss*, cuja métrica de IoU caiu de 0.9035 (com  $10^{-4}$ ) para 0.5577. Esse comportamento aponta para dificuldades de convergência e instabilidade no processo de otimização com taxas muito elevadas.

Embora funções como a *Cross Entropy* e a *Jaccard Loss* tenham mantido desempenho relativamente alto mesmo com  $10^{-2}$ , os resultados globais indicam que essa taxa não é ideal para a maioria das funções testadas.

Dessa forma, conclui-se que *learning rates* de  $10^{-4}$  ou  $10^{-3}$  são mais apropriadas para esta tarefa, proporcionando um equilíbrio entre desempenho e estabilidade. A interação entre taxa de aprendizagem e função de perda mostrou-se crítica, reforçando a necessidade de ajuste conjunto desses parâmetros durante a fase de experimentação.

### 4.3 Análise Visual da Segmentação

Para complementar a avaliação quantitativa, foi realizada uma análise visual da segmentação produzida pelo modelo. A Figura 9 apresenta um exemplo ilustrativo contendo: (a) a imagem original; (b) a máscara de segmentação de referência (*ground truth*); (c) a máscara prevista pelo modelo; e (d) a sobreposição dos erros de predição, destacando os falsos positivos e falsos negativos.

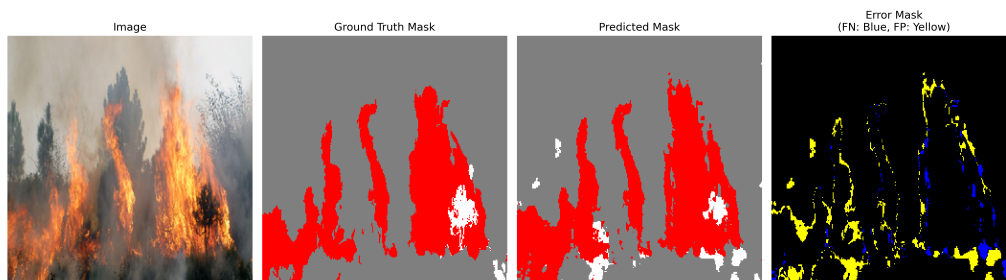


Figure 9: Análise visual da segmentação: (a) imagem original; (b) máscara de referência; (c) predição do modelo; (d) mapa de erro com falsos positivos (a amarelo) e falsos negativos (em azul).

Na imagem original (a), observa-se a presença de regiões com fogo e/ou fumo, as quais foram rotuladas manualmente na máscara de referência (b). A predição do modelo (c) mostra uma segmentação bastante próxima da verdade de terreno, evidenciando a boa capacidade do modelo em identificar corretamente as regiões relevantes.

Na imagem (d), é apresentado um mapa de erro, onde os **falsos positivos** (pixels identificados como fogo/fumo pelo modelo, mas ausentes na máscara real) são assinalados a amarelo, e os **falsos negativos** (pixels presentes na máscara real, mas ignorados pela predição) são assinalados a azul. Essa visualização permite identificar de forma clara os padrões de erro do modelo.

Observa-se que a maioria dos erros está concentrada em regiões de contorno, o que é típico em tarefas de segmentação semântica. A baixa ocorrência de falsos negativos sugere que o modelo é conservador na detecção de regiões críticas, o que é vantajoso em aplicações de segurança e monitoramento. Por outro lado, os falsos positivos, embora presentes, são dispersos e não formam regiões coesas, o que indica boa precisão geral.

Esta análise qualitativa valida os resultados quantitativos obtidos nas métricas, reforçando a capacidade do modelo em realizar segmentações precisas mesmo em cenários com complexidade visual elevada.

## 5 - Calendário

O cronograma apresentado na figura 10 detalha as etapas principais do projeto, distribuídas entre os meses de outubro de 2024 e junho de 2025. Este plano de trabalho organiza-se em cinco atividades principais: pesquisa científica, construção do dataset, desenvolvimento do algoritmo U-Net, desenvolvimento do algoritmo ResUNet-a e redação do relatório. A seguir, é apresentada uma descrição detalhada de cada uma das fases.

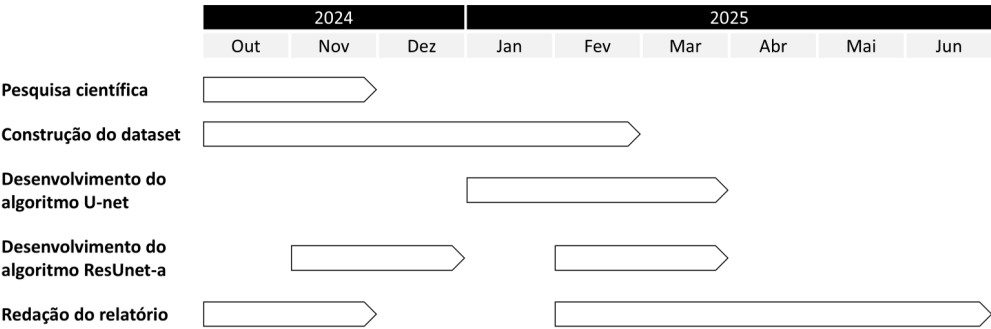


Figure 10: Calendário de atividades do TFC

### Pesquisa Científica

Esta etapa inicial foca na revisão de literatura e exploração de metodologias já existentes para a segmentação de incêndios florestais. Serão analisados artigos científicos, algoritmos de aprendizagem profunda (como U-Net e ResUNet-a) e estudos que visam identificar as métricas mais relevantes para avaliar a eficácia de modelos de segmentação. O objetivo principal desta fase é construir uma base teórica sólida para guiar o desenvolvimento técnico do projeto.

### Construção do Dataset

Durante este período, será realizada a construção de um dataset personalizado para a segmentação de incêndios florestais, utilizando imagens dos conjuntos de dados Corsican Dataset e Firefront Gestosa Dataset. Serão criadas manualmente máscaras detalhadas para identificar os pixels de fogo, fumo e fundo nas imagens, com o objetivo de garantir a precisão e a qualidade dos dados para o treino dos modelos.

### Desenvolvimento do Algoritmo U-Net

Uma vez construído o dataset, inicia-se o desenvolvimento do modelo U-Net. Nesta fase, será implementado um modelo de segmentação que utiliza a arquitetura U-Net para identificar e segmentar áreas afetadas por incêndios. O processo inclui o treino do modelo, ajustes nos seus hiperparâmetros e a avaliação. O objetivo desta fase é criar um modelo robusto que consiga segmentar áreas de fogo e fumo com alta fiabilidade.

### Desenvolvimento do Algoritmo ResUNet-a

Durante este período, será desenvolvido o algoritmo ResUNet-a, uma versão avançada do modelo de segmentação, projetada para lidar com tarefas complexas, como a seg-

mentação de áreas afetadas por incêndios florestais. O ResUNet-a incorpora técnicas avançadas, como convoluções dilatadas (atrous convolutions) e mecanismos de atenção espacial (spatial attention mechanisms), que permitem ao modelo capturar padrões espaciais e contextuais mais ricos em imagens de alta resolução. O foco será melhorar a precisão da segmentação em condições desafiadoras. Durante esta fase, o modelo será treinado, ajustado e avaliado, com especial atenção aos hiperparâmetros, visando otimizar seu desempenho e garantir sua capacidade de generalização para novos dados.

## **Redação do Relatório**

A redação do relatório ocorrerá ao longo de todo o projeto, com maior intensidade nos meses finais. Esta atividade inclui a documentação detalhada das etapas do projeto, desde a metodologia e construção do dataset até os resultados obtidos nos testes com os modelos. A entrega de um relatório bem estruturado e completo é essencial para consolidar os resultados do projeto e garantir que as conclusões sejam devidamente comunicadas.

Este cronograma foi elaborado para permitir uma progressão lógica das atividades, assegurando que cada etapa seja concluída com a profundidade necessária para alcançar os objetivos estabelecidos no projeto.

## 6 - Conclusão

### 6.1 Síntese dos Resultados

Este trabalho teve como objetivo desenvolver e avaliar modelos de segmentação semântica para a identificação de fogo e fumo em imagens de incêncios florestais, com base nas arquiteturas U-Net e ResUNet. A análise foi conduzida com um conjunto reduzido de dados anotados manualmente — um total de 375 imagens — o que, embora limitado, permitiu alcançar resultados promissores.

Inicialmente, utilizou-se a U-Net como baseline com a função de perda Cross Entropy, observando-se um bom desempenho, especialmente com *learning rate* de  $10^{-4}$  e  $10^{-3}$ . Em seguida, a arquitetura ResUNet foi avaliada de forma mais abrangente, testando cinco funções de perda e três taxas de aprendizagem. Os melhores resultados foram obtidos com Cross Entropy e Jaccard Loss, particularmente quando combinadas com *learning rates* mais baixos ( $10^{-4}$  e  $10^{-3}$ ), demonstrando elevada *IoU* e consistência nas restantes métricas.

### 6.2 Importância da Taxa de Aprendizagem

A taxa de aprendizagem mostrou-se um dos fatores mais críticos para o desempenho dos modelos. Aprendizagens muito elevadas ( $10^{-2}$ ) resultaram frequentemente em instabilidade e degradação de desempenho, especialmente para funções de perda mais sensíveis como a Tversky Loss. Por outro lado, taxas menores ( $10^{-4}$  e  $10^{-3}$ ) proporcionaram uma aprendizagem mais estável, com convergência mais eficaz e resultados mais consistentes. Este achado reforça a importância de uma escolha criteriosa do *learning rate* durante a fase de ajuste de hiperparâmetros.

### 6.3 Impacto do Conjunto de Dados e Possíveis Melhorias

O conjunto de dados utilizado constitui uma limitação significativa. Apesar dos bons resultados obtidos, a utilização de apenas 375 imagens segmentadas manualmente restringe a capacidade de generalização do modelo para diferentes cenários e condições visuais. Ampliar o conjunto de dados, seja por meio da anotação manual de novas imagens, seja por técnicas de *data augmentation* (como rotações, reflexões, alterações de brilho e contraste), poderá melhorar significativamente a robustez e o desempenho do modelo.

### 6.3 Impacto do Conjunto de Dados e Possíveis Melhorias

O conjunto de dados utilizado constitui uma limitação significativa. Apesar dos bons resultados obtidos, a utilização de apenas 375 imagens segmentadas manualmente restringe a capacidade de generalização do modelo para diferentes cenários e condições visuais. Além disso, por se tratar de anotações manuais, é possível que existam erros ou imprecisões nas máscaras de referência, o que pode introduzir ruído no processo de treino e avaliação.

Ampliar o conjunto de dados, seja por meio da anotação de novas imagens com maior rigor, seja pela adoção de técnicas de *data augmentation* (como rotações, reflexões, al-

terações de brilho e contraste), poderá melhorar significativamente a robustez e o desempenho do modelo.

## **6.4 Considerações Finais**

Os resultados obtidos demonstram que a ResUNet, combinada com funções de perda adequadas e uma taxa de aprendizagem bem ajustada, é capaz de realizar segmentações precisas de fogo e fumo, mesmo em contextos visuais complexos. A análise quantitativa foi reforçada por uma avaliação visual das segmentações, a qual revelou uma boa correspondência entre as previsões do modelo e as máscaras de referência, com erros principalmente localizados nas regiões de contorno.

Embora promissora, a aplicação prática deste modelo requer novos testes com dados mais variados e em tempo real, além da integração em sistemas de monitoramento e detecção automática. Este trabalho oferece uma base sólida sobre a qual futuras investigações poderão construir sistemas mais completos, robustos e aplicáveis a cenários reais de risco.



# Bibliografia

- [1] *Papers with Code - U-Net Explained*. URL: <https://paperswithcode.com/method/u-net>.
- [2] Giuliana Zanchi et al. "Simulation of water and chemical transport of chloride from the forest ecosystem to the stream". In: *Environmental Modelling & Software* (2021). URL: <https://www.sciencedirect.com/science/article/pii/S136481522100027X>.
- [3] Jesús San-Miguel-Ayanz, Jose Manuel Moreno, and Andrea Camia. "Analysis of large fires in European Mediterranean landscapes: Lessons learned and perspectives". In: *Forest Ecology and Management* (2013). URL: <https://www.sciencedirect.com/science/article/pii/S0378112712006561>.
- [4] Tom Toulouse et al. *Computer vision for wildfire research: An evolving image dataset for processing and analysis*. July 07, 2017. URL: <https://www.sciencedirect.com/science/article/abs/pii/S0379711217302114?via%3Dihub> (visited on 11/2024).
- [5] *Image Labeler*. URL: <https://www.mathworks.com/help/vision/ug/get-started-with-the-image-labeler.html>.
- [6] Guan et al. "Forest Fire Segmentation from Aerial Imagery Data Using an Improved Instance Segmentation Model". In: *Remote Sensing* 14.13 (2022). ISSN: 2072-4292. DOI: 10.3390/rs14133159. URL: <https://www.mdpi.com/2072-4292/14/13/3159>.
- [7] Songbin Li et al. *Optimized Deep Learning Model for Fire Semantic Segmentation*. April 21, 2022. URL: <https://www.techscience.com/cmc/v72n3/47492/html>.
- [8] Songbin Li et al. *Aerial imagery pile burn detection using deep learning: The FLAME dataset*. March 23, 2021. URL: <https://www.sciencedirect.com/science/article/pii/S1389128621001201>.
- [9] *EFFIS*. URL: <https://forest-fire.emergency.copernicus.eu/>.
- [10] *Ciclope*. URL: <https://inov.pt/project/ciclope/index.html>.
- [11] Foivos I. Diakogiannis et al. "ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data". In: *ISPRS Journal of Photogrammetry and Remote Sensing* (2020). DOI: <https://doi.org/10.1016/j.isprsjprs.2020.01.013>. URL: <https://www.sciencedirect.com/science/article/pii/S0924271620300149>.
- [12] *Global<sub>Accuracy</sub>*. URL: [https://scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy\\_score.html](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy_score.html).
- [13] *Balanced<sub>Accuracy</sub>*. URL: [https://scikit-learn.org/stable/modules/generated/sklearn.metrics.balanced\\_accuracy\\_score.html](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.balanced_accuracy_score.html).
- [14] *IOU*. URL: [https://scikit-learn.org/stable/modules/generated/sklearn.metrics.jaccard\\_score.html](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.jaccard_score.html).
- [15] *F1<sub>score</sub>*. URL: [https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1\\_score.html](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html).
- [16] *Cross<sub>EntropyLoss</sub>*. URL: <https://pytorch.org/docs/stable/generated/torch.nn.CrossEntropyLoss.html>.

- [17] *Dice<sub>Tversky</sub>Jaccard<sub>Loss</sub>*. URL: <https://docs.monai.io/en/stable/losses.html>.
- [18] *GeneralizedDiceLoss*. URL: <https://docs.monai.io/en/stable/losses.html>.