|  |  |
| --- | --- |
|  | UNIVERSIDADE ESTADUAL DO NORTE DO PARANÁ  *Campus* LUIZ MENEGHEL - centro de ciências Tecnológicas  Ciência DA COMPUTAÇÃO |

WILLIAM SDAYLE MARINS SILVA

DETECÇÃO DE FOGO EM FRAMES DE VÍDEOS UTILIZANDO APRENDIZAGEM DE MÁQUINA

Bandeirantes

2018

William sdayle marins silva

detecção de fogo em frames de vídeos utilizando aprendizagem de máquina

Trabalho de Conclusão de Curso submetido à Universidade Estadual do Norte do Paraná, como requisito parcial para obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Bruno M. N. Souza

Bandeirantes

2018

William Sdayle marins silva

Detecção de fogo em frames de vídeo utilizando aprendizagem de máquina

Trabalho de Conclusão de Curso submetido à Universidade Estadual do Norte do Paraná, como requisito parcial para obtenção do grau de Bacharel em Ciência da Computação.

**COMISSÃO EXAMINADORA**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Prof. Me. Bruno Miguel Nogueira de Souza

UENP – *Campus* Luiz Meneghel

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Prof. Me. Glauco Carlos Silva

UENP – *Campus* Luiz Meneghel

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Prof. Dr. Ederson Marcos Sgarbi

UENP – *Campus* Luiz Meneghel

Bandeirantes, 26 de novembro de 2018.

Dedico este trabalho a meu pai, que foi a pessoa

que mais me apoiou no decorrer do curso.

**AGRADECIMENTOS**

Agradeço primeiramente a Deus por ter nos deixado chegar até aqui, posteriormente gostaria de agradecer aos meus pais, por sempre estarem ao meu lado, ajudando e me aconselhando de maneira extraordinária. Gostaria também de agradecer ao meu orientador mestre Bruno Miguel Nogueira de Souza, pelos momentos retirados para me ajudar fora de hora, pela maneira de ensino e sugestões, agradecer aos membros da banca, ao Me. Glauco Carlos Silva pelas sugestões relatadas nas apresentações feitas durante o ano, ao Dr. Ederson Marcos Sgarbi meus agradecimentos pela maneira sugerida para apresentar o trabalho. Agradecer aos meus amigos pelos momentos de alegria durante os anos, em especial aos amigos da turma: Junior, André, Felipe e Rafael e os demais que iniciam a jornada ao meu lado.

Nunca diga nunca, pois os limites são como os medos: sempre são apenas ilusão. (Michael Jordan)

**Resumo**

Incêndios podem ocorrer em qualquer momento, basta haver os objetos necessários para que isso ocorra, sendo assim podem ocorrer em qualquer local, deixando vários problemas por onde passa. Partindo dessa ideia, este trabalho foi proposto para realização de uma detecção de fogo em vídeos, utilizando redes neurais convolucionais para validação das imagens. Foi levantada uma base de vídeos, onde as imagens possuíam características semelhantes, as de conter fogo e fumaça, a partir disso foram gerados frames contendo apenas o movimento que foi detectado no vídeo. Foram utilizados métodos para remoção de fundo baseado em misturas gaussianas e morfologia matemática, após a extração dos fundos, as imagens continham apenas as áreas de interesse, sendo assim as imagens foram utilizadas em redes neurais convolucionais, as arquiteturas utilizadas foram Xception e ResNet. Os resultados apresentados por ambas as redes, foram extremamente satisfatórios com o propósito do trabalho.

**Palavras chave:** Fogo, vídeos e redes neurais convolucionais.

**Abstract**

Fires can occur at any time, just having the objects needed for it, so they can occur anywhere leaving several problems wherever it goes. Based on this idea, this work was proposed to perform a fire detection in videos, using convolutional neural networks to validate the image of the videos. A base with videos was created, containing videos with similar characteristics, those of fire and smoke, from which were generated images containing only the movement that was detected in the video. After the extraction of the backgrounds, the images only contained parts of interest, after that, the images were used in convolutional neural networks, which the architectures were the Xception and ResNet. The results presented by both networks were extremely satisfactory for the propose of this work.

**Key-words:** Fire, videos and convolutional neural networks.

LISTA DE FIGURAS

[Figura 1: Valores de incêndios em 32 países – Fonte: Corrêa & Braga, 2007. 19](#_Toc530757429)

[Figura 2: Valores de mortes relacionadas a incêndios - Fonte: Corrêa & Braga, 2007. 19](#_Toc530757430)

[Figura 3: Edifício Andraus em chamas - Fonte: Seito et al., 2008 20](#_Toc530757431)

[Figura 4: Hierarquia para processamento de imagens - Fonte: Eustáquio, Queiroz, & Gomes, 2001 21](#_Toc530757432)

[Figura 5: Ilustração de como são passados os *frame*s nos vídeos - Fonte: Sonaglio & Moecke, 2008. 23](#_Toc530757433)

[Figura 6: Erosão de conjuntos pretos a) Imagem original, b) Imagem aplicado elemento estruturante , c) Imagem com elemento estruturante 3x . 24](#_Toc530757434)

[Figura 7: Dilatação de conjuntos pretos. a) Imagem original, b) Imagem aplicado elemento estruturante , c) Imagem com elemento estruturante 3x . 25](#_Toc530757435)

[Figura 8: Áreas onde pode ser aplicado inteligência artificial - Fonte: Dos & Gomes, 2010. 27](#_Toc530757436)

[Figura 9: Formas de aprendizado - Fonte: Ferauche & Almeida, 2011. 28](#_Toc530757437)

[Figura 10: Neurônio biológico - Fonte: Barreto (2002) 29](#_Toc530757438)

[Figura 11: Modelo de circuito binário para um neurônio artificial (*perceptron*) – Fonte: Rauber (2010) 30](#_Toc530757439)

[Figura 12: Funções de ativação - Fonte: Rauber, 2010 31](#_Toc530757440)

[Figura 13: Rede neural simplificada. Fonte: Ferneda (2006) 31](#_Toc530757441)

[Figura 14: Etapas para utilização em redes neurais - Fonte: Zeiler & Fergus, 2012. 32](#_Toc530757442)

[Figura 15: Ilustração das etapas de aprendizagem - Fonte: Goodfellow, Ian, Yoshua Bengio, 2016. 33](#_Toc530757443)

[Figura 16: Arquitetura do módulo *Inception* - Fonte: Ushizima (2017) 35](#_Toc530757444)

[Figura 17: Arquitetura *Xception*, onde os dados passam pelo fluxo de entrada, posteriormente pelo fluxo médio e por fim passa pelo fluxo de saída. Fonte: Chollet (2014) 36](#_Toc530757445)

[Figura 18: Demonstração de como é entrado cada valor nas camadas da rede ResNet– Fonte: He & Sun (2015) 37](#_Toc530757446)

[Figura 19: Matriz de confusão - Fonte: Fawcett (2006) 38](#_Toc530757447)

[Figura 20: Metodologia adotada para realização do trabalho 43](#_Toc530757448)

[Figura 21: *Frames* dos vídeos selecionados para o trabalho 45](#_Toc530757449)

[Figura 22: Imagens dos *frame*s onde foi detectado movimento 46](#_Toc530757450)

[Figura 23: Código utilizado para gravar imagens já processadas 46](#_Toc530757451)

[Figura 24: Imagens já processadas e apenas com movimento detectado 47](#_Toc530757452)

[Figura 25: Mostrando o resultado obtido após a utilização do método GB (*Gaussian Blur*). 50](#_Toc530757453)

[Figura 26: Resultado após a utilização do método para diferença absoluta 50](#_Toc530757454)

[Figura 27: Imagem utilizada como máscara para detecção de movimento 51](#_Toc530757455)

[Figura 28: Imagem final após passar pelo processo de detecção de movimento e remoção de *background* 51](#_Toc530757456)

[Figura 29: Acurácia em relação a iterações na base de treino utilizando os *frame*s de vídeos 52](#_Toc530757457)

[Figura 30: F-Score em relação a iterações na base de treino utilizando os *frame*s de vídeos 53](#_Toc530757458)

[Figura 31: Log *Los*s em relação a iterações na base de treino utilizando os *frame*s de vídeos 53](#_Toc530757459)

[Figura 32: Erro médio absoluto em relação a iterações na base de treino utilizando os *frame*s de vídeos 54](#_Toc530757460)

[Figura 33: Acurácia após ser aplicado *frame*s na base de validação 55](#_Toc530757461)

[Figura 34: F-Score após ser aplicado os *frame*s na base de validação 56](#_Toc530757462)

[Figura 35: *Log Loss* após ser utilizado os *frame*s na base de validação 56](#_Toc530757463)

[Figura 36: Erro Médio Absoluto após ser aplicado os *frame*s na base de validação 57](#_Toc530757464)

[Figura 37: Acurácia em relação a iterações na base de validação utilizando os *frame*s de vídeos na rede ResNet 58](#_Toc530757465)

[Figura 38: F-Score em relação a iterações na base de validação utilizando os *frame*s de vídeos na rede ResNet 59](#_Toc530757466)

[Figura 39: *Log Loss* em relação a iterações na base de validação utilizando os *frame*s de vídeos na rede *ResNet.* 59](#_Toc530757467)

[Figura 40: Erro Médio Absoluto em relação a iterações na base de validação utilizando os *frame*s de vídeos na rede *ResNet.* 60](#_Toc530757468)

LISTA DE TABELAS

Tabela 1: Valores recolhidos pela rede Xception na base de teste......................................................54

Tabela 2: Melhores valores recolhidos pela rede neural Xception, na base de validação...................57

Tabela 3:Melhores valores recolhidos pela rede ResNet....................................................................60

SUMÁRIO

[1. Introdução 14](#_Toc530583726)

[1.1 Contexto e Delimitação do Trabalho 15](#_Toc530583727)

[1.2 Formulação do Problema 16](#_Toc530583728)

[1.3 Objetivos 16](#_Toc530583729)

[1.3.1 Objetivo Geral 16](#_Toc530583730)

[1.3.2 Objetivos Específicos 16](#_Toc530583731)

[1.4 Justificativa 17](#_Toc530583732)

[1.5 Organização do Trabalho 17](#_Toc530583733)

[2. Fundamentação Teórica 18](#_Toc530583734)

[2.1 Fogo e seus impactos na sociedade 18](#_Toc530583735)

[2.2 Processamento Digital de Imagens. 21](#_Toc530583736)

[2.2.1 Imagem Digital 22](#_Toc530583737)

[2.2.2 Morfologia Matemática 24](#_Toc530583738)

[2.3 Detecção de Movimento 25](#_Toc530583739)

[2.3.1 Remoção de fundo em vídeos 25](#_Toc530583740)

[2.4 Inteligência Artificial 26](#_Toc530583741)

[2.5 Métricas de Avaliação 38](#_Toc530583742)

[2.6 Trabalhos relacionados a detecção de fogo 40](#_Toc530583743)

[3. Metodologia 43](#_Toc530583744)

[3.1 Base de Imagens a partir de vídeos 44](#_Toc530583745)

[3.2 Detectando regiões de interesse 46](#_Toc530583746)

[3.3 Aplicação das imagens em redes neurais 47](#_Toc530583747)

[4. Desenvolvimento 49](#_Toc530583748)

[5. Conclusão 62](#_Toc530583749)

[6. Referencias 64](#_Toc530583750)

# Introdução

Fogo é um processo químico em constante transformação. Também pode ser definido como o resultado de uma reação química onde a luz e calor se separam devido à combustão de materiais diversos. Os elementos nos quais compõem o fogo são os seguintes: combustivel, comburente, calor e a reação em cadeia. O combustível pode ser elemento sólido como papel, madeira e tecido; elementos líquidos como gasolina, álcool e elementos gasosos como butano, propano e etano. O comburente ou oxigênio é o componente que incita o fogo de acordo com SP (2010) que ao se combinar com os vapores inflamáveis dos combustíveis, desta forma ativa as chamas e confere maior possibilidade de que ela se alastre.

O fogo possui um papel importante na sociedade, uma vez que por meio de sua descoberta, foi possível melhorar a qualidade da vida do ser humano, ou seja, aquecendo ambientes, mudando a forma da alimentação, criando mecanismos de proteção, locomoção entre outros aspectos. Portanto, a domesticação do fogo foi benéfica para a evolução da espécie humana. Porém, o fogo não controlado, fomenta riscos ao ambiente ecológico e a vida humana.

De acordo com Castro et al. (2003), existem 3 principais fatores para incêndios florestais, sendo eles: características dos combustíveis, característica do relevo e condições meteorológicas, afirmando que o incêndio começa em um local menor e se alastra por encontrar novos combustíveis.

Visando o combate ao incêndio de forma precoce, a tecnologia pode auxiliar neste processo. Dentre os mecanismos tecnológicos para este fim, podemos citar a utilização de sensores de fumaça e calor, porém, estes equipamentos, devem estar próximos do local afetado. Neste sentido, uma outra forma de localizar o foco de incêndio é utilizando o monitoramento remoto em vídeo de uma região com alta incidência ou probabilidade de acontecer incêndios. Como avanço da área do processamento de imagens, é possível automatizar o processo de detecção de incêndios em vídeo, realizando o processamento das imagens de vídeo. Porém, esta não é uma tarefa trivial, pois requer a identificação de um elemento que não possui forma e nem tamanho pré-definido, além de possuir uma infinidade de variações de cores.

Baseado nos problemas de incêndio e nos desafios da detecção automática de fogo apresentados, a proposta deste trabalho é elaborar um mecanismo que seja capaz de detectar fogo em vídeos, com o objetivo de monitorar as áreas onde é possível ocorrer incêndios. Neste sentido, a proposta consiste em apresentar um modelo de detecção de fogo através câmeras de segurança para monitoramento de ambiente utilizando o processamento de imagens de vídeos afim de identificar o movimento e consequentemente realizar a detecção de fogo nas imagens.

## Contexto e Delimitação do Trabalho

Para que seja possível a detecção precoce de fogo por meio do processamento de vídeos, a área da Ciência da Computação chamada de Visão Computacional é utilizada neste trabalho. Neste contexto, o processamento digital de imagens, aborda a ideia de se ter uma imagem como entrada e que após algum processamento, será gerada uma nova imagem a partir das características, do processo envolvido (SPRING,1996).

Além do processamento de imagens, a aprendizagem de máquina também será utilizada como auxiliar na identificação das regiões de fogo. Dentro da área da aprendizagem de máquina as Redes Neurais Convolucionais (RNCs) (aprendizagem profunda) têm apresentado resultados satisfatórios para a identificação de objetos, neste sentido, tal abordagem será testada no que tange a identificação de fogo nas imagens pré-selecionadas.

## Formulação do Problema

Aplicar pré-processamento em vídeos, por meio de destaque das partes em movimento, e utilizar as imagens resultantes para treinar e avaliar modelos de aprendizagem de máquina podem otimizar o processo de detecção de fogo com baixa taxa erros e com alta acurácia.

## Objetivos

### Objetivo Geral

Trabalho tem como objetivo a prevenção de incêndios em locais que possuem câmeras de segurança, usando as mesmas para visualizar imagens captadas pela câmera, através de meios da computação gráfica para processar tais imagens, como: métodos para selecionar fundo de imagem e frente de imagem, separação dos *frame*s de um vídeo, detecção de movimento contido nas imagens captadas pelas câmeras, utilizar aprendizagem profunda para então realizar o reconhecimento de fogo dentro das imagens.

### Objetivos Específicos

* Realizar processamento de vídeos;
* Extrair *frame*s dos vídeos em formato de imagem;
* Utilizar imagens extraídas e realizar a comparação entre elas;
* Detectar movimento;
* Separar imagens que apresentem mudanças (Remoção de Fundo);
* Aplicar imagens em uma base;
* Utilizar redes neurais convolucionais para detecção de chamas.

## Justificativa

Além de Auxiliar na detecção precoce do incêndio e fumaça, o presente trabalho visa estudar as variações e aplicações computacionais de métodos de aprendizagem de máquina aplicados a imagens de vídeos pré-processadas.

## Organização do Trabalho

Na seção 2 é apresentado a fundamentação teórica, subseção 2.1 é apresentado a ideia de que é processamento de imagem, subseção 2.2 é apresentado o que é processamento de vídeo, subseção 2.3 apresentado aprendizagem de máquina, subseção 2.3.1 apresentado o deep learning, subseção 2.4 apresentado técnicas sobre remoção de *background*, seção 3 é apresentada a metodologia.

# Fundamentação Teórica

Neste capítulo serão apresentadas as teorias que envolvem o desenvolvimento deste trabalho. Neste sentido, na seção 2.1 será abordado sobre os impactos que incêndios não controlados trazem para a sociedade, já na seção 2.2 será tratado do tema de Processamento digital de Imagens, na seção 2.3 será detalhado o assunto sobre detecção de movimento, posteriormente na seção 2.4 serão abordados assuntos de aprendizagem de máquina, *deep learning* e redes neurais convolucionais, na seção 2.5 é onde serão mostrados as métricas de avaliação utilizadas para o trabalho e na seção 2.6 são trabalhos relacionados ao tema da detecção de fogo em imagens/vídeos.

## Fogo e seus impactos na sociedade

Existem 4 classes para um incêndio, conforme afirmado por SP (2010): classe A caracteriza-se por fogo em materiais sólidos, classe B em combustíveis líquidos inflamáveis, classe C em materiais energizados e classe D em metais pirofóricos como alumínio e magnésio.

Segundo Tamiozzo et al. (2017), os incêndios florestais geram problemas em diversas escalas. Localmente, promovem a degradação da vegetação, impactos sobre a biodiversidade, prejuízos financeiros e até mesmo perda de vidas, as características dos incêndios são afetadas pelas complexas interações entre vegetação, clima, topografia e atividades antrópicas ao longo do tempo.

De acordo com Corrêa & Braga (2007), incêndios em edificações (comerciais, de prestação de serviço, industriais ou residenciais) são frequentes em todo mundo. Estima-se que este tipo de incêndio represente um terço (1/3) de todas as ocorrências registradas. A figura 1 representada a baixo, expõe os dados em porcentagem e corrobora com a afirmação após levantamento realizado em 32 países.

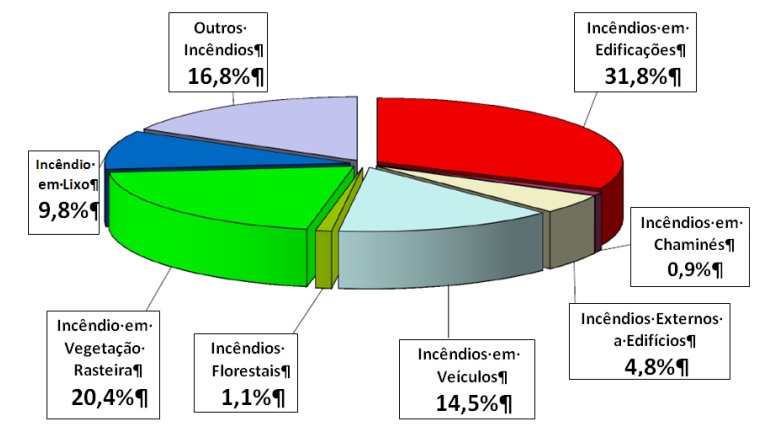


Figura 1: Valores de incêndios em 32 países – Fonte: Corrêa & Braga, 2007.

Corrêa & Braga (2007) apresenta números de mortes relacionadas com incêndios por grupo local registrado na Grã-Bretanha (UK) de acordo com as estatísticas de incêndio do Departamento de Comunidades e Governo Local dos anos de 2000 até 2012. Corrêa & Braga (2007) afirma também que a maior parte das mortes foram em habitações queimadas.

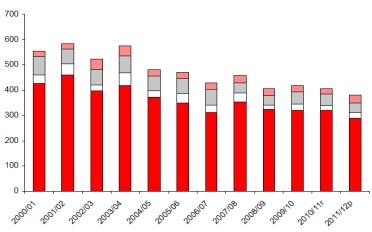


Figura 2: Valores de mortes relacionadas a incêndios - Fonte: Corrêa & Braga, 2007.

Veículos rodoviarios

Outdoors

Habitações

Outros edifícios  
outros edifícios

No Brasil, as causas dos incêndios florestais nas Unidades de Conservação têm sido principalmente devido ao uso incorreto do fogo por pessoas, para renovação de pastagens e limpeza de restos de cultura nas propriedades vizinhas. Além disso, são também causas frequentes a ação de incendiários, caçadores, pescadores e soltura de balões (Fiedler, 2003).

De acordo com Luiz, Tebaldi, Fiedler, Juvanhol, & Dias (2013), os equipamentos, ferramentas e treinamentos exigidos para prevenção e combate dos incêndios florestais são diferentes dos utilizados no combate aos incêndios urbanos, uma vez que apresentam especificidades que se adequem á diferentes situações. Afirma também que o fogo é o agente com grande potencial para modificar os ecossistemas naturais. Seu regime pode ser caracterizado pelo grau de alteração no ambiente, variável quanto á intensidade, da duração, da frequência e da vulnerabilidade da área afetada pelos incêndios.

O primeiro grande incêndio em prédios elevados ocorreu em 24 de fevereiro de 1972, no edifício Andraus, na cidade de São Paulo. Tratava-se de um edifício comercial e de serviços (Loja Pirani e escritórios), situado na Avenida São João esquina com Rua Pedro Américo, com 31 andares, estrutura em concreto armado e acabamento em pele de vidro. Acredita-se que o fogo tenha começado nos cartazes de publicidade das Casas Pirani, colocados sobre a marquise do prédio. (Seito et al., 2008)



Figura 3: Edifício Andraus em chamas - Fonte: Seito et al., 2008

## Processamento Digital de Imagens.

O Processamento Digital de Imagens (PDI) não é uma tarefa simples, na realidade envolve um conjunto de tarefas interconectadas. Tudo se inicia com a captura de uma imagem, a qual, normalmente, corresponde à iluminação que é refletida na superfície dos objetos, realizada através e um sistema de aquisição. Após a captura por um processo de digitalização, uma imagem precisa ser representada de forma apropriada para tratamento computacional. Imagens podem ser representadas em duas ou mais dimensões (Eustáquio, Queiroz, & Gomes, 2001).

A Figura 4 mostra como seria uma hierarquia para um processamento de imagens.

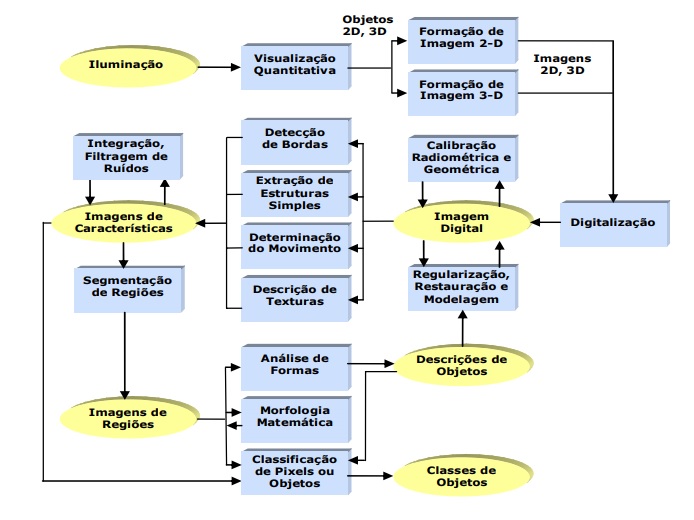


Figura 4: Hierarquia para processamento de imagens - Fonte: Eustáquio, Queiroz, & Gomes, 2001

Carlin et al. 2004 afirma que primeiro conceito é que qualquer imagem (tela de TV, figura, projeção, etc...) pode ser descrita por uma distribuição de fluxo bidimensional, a qual pode ser descrita pelo valor da irradiância I(y, z) que é uma função que possui um valor para cada ponto no espaço sobre o qual a imagem está distribuída.

Eustáquio, Queiroz, & Gomes (2001) faz uma comparação entre computação gráfica e processamento de imagens afirmando que em uma visão simplificada, computação gráfica busca imagens fotos-realísticas de cenas tridimensionais geradas por computador, enquanto processamento de imagem tenta reconstruir uma cena tridimensional a partir de uma imagem real, obtida através de uma câmera. Neste sentido, processamento de imagem busca um procedimento inverso ao de computação gráfica, análise ao invés de síntese, mas ambas as áreas atuam sobre o mesmo conhecimento, o qual inclui, dentre outros aspectos, a interação entre iluminação e objetos e projeções de uma cena tridimensional em um plano de imagem. O cenário envolvendo todas as disciplinas que tenham algum ingrediente de processamento da informação visual, dentre as quais a computação gráfica e o processamento de imagem ocupam posição de destaque, é definido por alguns autores como Computação Visual.

### Imagem Digital

Uma imagem monocromática é uma função bidimensional contínua f(x,y), na qual x e y são coordenadas espaciais e o valor de f em qualquer ponto (x,y) é proporcional à intensidade luminosa (brilho ou nível de cinza) no ponto considerado. Como os computadores não são capazes de processar imagens contínuas, mas apenas arrays de números digitais, é necessário representar imagens como arranjos bidimensionais de pontos (Eustáquio, Queiroz, & Gomes, 2001).

Na visão de Scuri (2002) em processamento de imagens, usa-se o modelo matricial e a computação gráfica geralmente se baseia no modelo de objetos vetoriais. Neste modelo os objetos são armazenados apenas a partir da descrição das coordenadas de seus vértices, sejam elas espaciais ou planares (três ou duas dimensões, respectivamente). Dessa maneira utiliza-se um sistema de coordenadas cartesiano, onde os objetos podem ser escalados, rotacionados e transladados com maior liberdade para cada objeto. O modelo matricial utiliza uma matriz de dados para armazenar a informação de cor em cada ponto da imagem, onde o sistema de coordenadas é obviamente uma grade de números inteiros que descrevem a posição na matriz.

No processamento digital de sinais, uma imagem digital estática é representada por uma matriz de píxeis. Esta matriz pode ser bidimensional, representando apenas a luminância (preto e branco), ou tridimensional (imagens coloridas). Cada plano bidimensional de uma imagem colorida representa uma das cores do modelo de cor aditivo *RGB* (vermelho, verde e azul), ou um dos componentes dos espaços de cores, como por exemplo o *YCbCr*. Na imagem, cada pixel *RGB* é normalmente representado por 24 bits, ou seja, 8 bits para cada componente de cor, sendo que o valor do pixel em cada componente *RGB* pode variar entre 0 (preto) e 255 (branco). A resolução da imagem corresponde ao número de colunas e de linhas de píxeis contidas em sua matriz (Sonaglio & Moecke, 2008).

A Figura 5 ilustrará como são as dimensões dos canais de cor em quadros (*frames*) que são passados dentro dos vídeos.

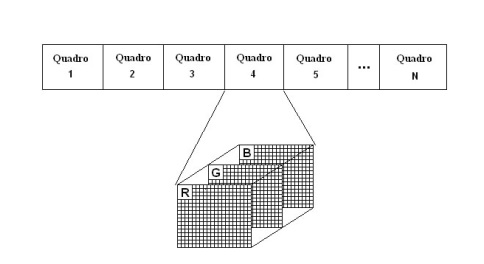


Figura 5: Ilustração de como são passados os *frame*s nos vídeos - Fonte: Sonaglio & Moecke, 2008.

De acordo com Jr (2015), a frequência espacial de uma imagem é dada pelas variações de intensidade por unidade e distancia, afirmando que filtragem espacial de frequência consiste de operadores espaciais que filtram determinadas frequências no sentido de suavizar. Filtros podem ser utilizados de forma computacional de duas maneiras: convolução ou análise de Fourier. Jr (2015) afirma que filtros gaussianos são utilizados para borrar ou desfocar a imagem, o resultado dessa operação é a suavização da imagem. Matematicamente a função gaussiana é realizada com um filtro passa baixa e outro passa alta.

### Morfologia Matemática

No trabalho de Facon (2011), foi afirmado que morfologia matemática consiste em extrair informações relativas a geometria e a topologia de conjuntos desconhecidos de uma imagem a partir de uma determinado elemento estruturante. Suas principais operações são erosão e dilatação. Erosão sendo possível diminuir geometricamente os objetos contidos nas imagens, dilatação sendo possível aumentar o tamanho dos objetos nas imagens, os dois processos utilizam algum tipo de elemento estruturante. O elemento estruturante pode variar de acordo com o formato geométrico do objeto que se encontra na imagem. Facon (2011) afirma que existe um número limitado de elementos estruturantes,é um exemplo de elemento estruturante, apresentado no formato de matriz 3x3.

A figura 6 mostra o resultado do processo de erosão a partir de um elemento estruturante 3x3.

Uma imagem contendo relógio, objeto

Descrição gerada automaticamente

Figura 6: Erosão de conjuntos pretos a) Imagem original, b) Imagem aplicado elemento estruturante , c) Imagem com elemento estruturante 3x .

Na Figura 7 é mostrado o resultado do processo de dilatação de objetos em uma imagem utilizando o mesmo elemento estruturante 3x3.

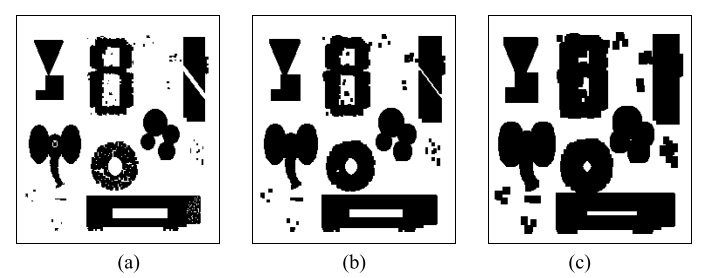


Figura 7: Dilatação de conjuntos pretos. a) Imagem original, b) Imagem aplicado elemento estruturante , c) Imagem com elemento estruturante 3x .

## Detecção de Movimento

Uma das maneiras plausíveis de se detectar movimento é através de sinais de vídeo, porém nesse método é necessário que a predominância de cena seja estática, ou seja, em ambientes que a câmera não se movimenta (Sonaglio & Moecke, 2008).

De acordo com Sonaglio & Moecke (2008), podem ser utilizados dois métodos para a detecção de movimento dentro dos *frame*s, sendo eles subtração do fundo para a detecção de movimento propriamente dita e a diferença temporal entre quadros sucessivos.

### Remoção de fundo em vídeos

O valor de um pixel no tempo t em *RGB* ou em algum outro espaço é denotado por x(t). A subtração de fundo baseada em pixels envolve decisão se o pixel pertencer ao *background* (BG) ou a algum objeto em primeiro plano (FG) (Zivkovic, 2004).

De acordo com Zivkovic (2004), os resultados da subtração de plano de fundo geralmente são propagados para alguns módulos de nível mais alto, por exemplo, os objetos detectados geralmente são rastreados. Ao rastrear um objeto, poderíamos obter algum conhecimento sobre a aparência do objeto rastreado e esse conhecimento poderia ser usado para melhorar a subtração de fundo.

#### Reconhecimento de fundo

De acordo com (Zivkovic, 2004) um pixel pertence ao *background* se:

|  |  |
| --- | --- |
|  | (2.1) |

Onde é um valor de limiar, sendo referenciado como modelo de *background*. O modelo de fundo é estimado a partir de um conjunto de treinamento indicado como X e assumir distribuição uniforme para a aparência do objeto em primeiro plano .

***GMM –* Modelo de Mistura Gaussiana(*Gaussian Mixture Model*)**

Para Chen, He, & Wang (2010) o primeiro passo para reconhecimento de *background* é usar um modelo de mistura gaussiana (GMM) para distinguir os objetos em movimento em primeiro plano do fundo ainda plano. O modelo de mistura gaussiana é um algoritmo de modelagem de fundo adaptativo eficaz e simples que rastreia a história de cada pixel com uma mistura de distribuições KGaussianas, respectivamente, em uma sequência de vídeo. De acordo com Steffens (2015), esse método se tornou muito conhecido devido a sua eficiência em modelar distribuições multi-modais do segundo plano, o que ocorre por exemplo como árvores balançando no fundo ou reflexo da luz.

## Inteligência Artificial

Conforme apresentado em Dos & Gomes (2010), a inteligência artificial é um ramo da Ciência da Computação cujo interesse é fazer com que os computadores pensem ou se comportem de forma inteligente. Por ser um tópico muito amplo, IA também está relacionada com psicologia, biologia, lógica matemática, linguística, engenharia, filosofia, entre outras áreas científicas, conforme apresentado na Figura 8:

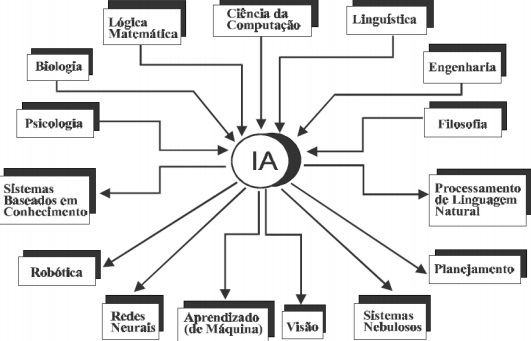


Figura 8: Áreas onde pode ser aplicado inteligência artificial - Fonte: Dos & Gomes, 2010.

A inteligência artificial é uma das ciências mais recentes, teve início após a Segunda Guerra Mundial e, atualmente, abrange uma enorme variedade de subcampos, desde áreas de uso geral, como aprendizado e percepção, até tarefas específicas como jogos de xadrez, demonstração de teoremas matemáticos, criação de poesia e diagnóstico de doenças (Dos & Gomes, 2010).

Dos & Gomes (2010) afirma que mostram uma linha de pensamento para se chegar a um conceito de o que é inteligência artificial:

1. Sistemas que pensam como seres humanos
2. Sistemas que atuam como seres humanos
3. Sistemas que atuam racionalmente

### Aprendizagem de Máquina

De acordo com Falciforme (2009), pode-se dizer que aprendizado de máquina é uma área da Inteligência Artificial que lida com problemas de aprendizado computacional a fim de adquirir conhecimento de forma automática. Um sistema de aprendizado tem a função de analisar informações e generalizá-las, para a extração de novos conhecimentos. Para a indução derivar conhecimento novo representativo, os exemplos das classes têm que estar bem-definidos e ter uma quantidade suficiente de exemplos, obtendo assim hipóteses úteis para um determinado tipo de problema. O aprendizado indutivo pode ser dividido em supervisionado e não-supervisionado, conforme apresentado na Figura 9:

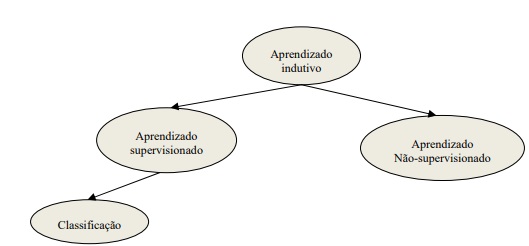


Figura 9: Formas de aprendizado - Fonte: Ferauche & Almeida, 2011.

Portanto, de acordo com Silva (2015), é possível definir aprendizado de máquina como área da inteligência artificial baseada em algoritmos que utilizam um grande número de exemplos para o treinamento de modelos computacionais.

### ****Redes neurais Convolucionais e Aprendizagem Profunda****

Segundo Silva (2015) *deep learning* trata-se de um conjunto de técnicas que utilizam redes neurais artificiais profundas, com muitas camadas intermediárias entre a camada de entrada e a de saída, afirmando também que o diferencial tecnológico dessa abordagem está nos excelentes resultados obtidos na resolução de problemas, resultados esses que superam até mesmo o desempenho dos melhores especialistas em determinadas áreas de conhecimento.

É importante considerar, também, o fato de que a estrutura hierárquica das camadas do córtex visual inspirou a criação de redes neurais convolucionais (Silva,2015).

**Neurônios**

Barreto (2002) afirma que, pode-se atribuir as principais funções cerebrais a os neurônios, afirmando que existem aproximadamente neurônios no cérebro humano. A figura 10 apresenta uma ilustração da forma de um neurônio.

Uma imagem contendo texto, mapa

Descrição gerada automaticamente

Figura 10: Neurônio biológico - Fonte: Barreto (2002)

Barreto (2002) afirma que o neurônio possui um corpo celular chamado de soma e mais algumas ramificações. Uma delas tem o nome de dendritos, eles são os responsáveis por conduzirem sinais das extremidades para o corpo celular, outra ramificação é chamada de axônio, sua função é transmitir o sinal do corpo celular para as extremidades, suas extremidades são conectadas com os dendritos de outros neurônios pelas sinapses. As sinapses possuem um papel fundamental no armazenamento de informação, afirma-se que um neurônio tem de 1000 a 10000 sinapses e pode receber informações compartilhadas com 1000 neurônios.

Durante alguns anos foram estudadas maneiras de representação de um neurônio artificial, até que Warren S. McCulloch criou um modelo de funcionamento de um neurônio em um circuito binário (Rauber,2010).

Uma imagem contendo texto

Descrição gerada automaticamente

Figura 11: Modelo de circuito binário para um neurônio artificial (*perceptron*) – Fonte: Rauber (2010)

As entradas para os neurônios são binárias, segundo Rauber (2010), cada entrada está associada a um peso , refletindo a importância da entrada , afirmando que o resultado dessa combinação linear é o valor da rede, esse resultado é dado a partir da equação 2.2, o resultado gerado pode ser chamado de combinação linear.

|  |  |
| --- | --- |
|  | (2.2) |
|  |  |

O valor do resultado, chamado como rede, serve como argumento para uma função chamada de função de transferência, gerando a resposta do neurônio (Barreto, 2002). Caso o valor ultrapasse um limiar µ, o neurônio dispara o valor 1 para saída binária y. O processo de comparação do resultado com o limiar é chamado de função de ativação, podendo ser realizada de 3 maneiras (Rauber,2010).

|  |  |  |
| --- | --- | --- |
| Função escada | Função linear | Função sigmoidal |
| Θ  Θ = 0 | Θ  Θ = 0  Θ |  |

Uma imagem contendo objeto, antena, céu

Descrição gerada automaticamente

Figura 12: Funções de ativação - Fonte: Rauber, 2010

O resultado final, chamado como y na figura 11, é chegado a partir da equação 2.3.

|  |  |
| --- | --- |
|  | (2.3) |

Redes Neurais Convolucionais

De acordo com Ferneda (2006), a habilidade de um ser humano em realizar funções complexas e principalmente a sua capacidade de aprender advêm do processamento paralelo e distribuído da rede de neurônios do cérebro. Os neurônios do córtex, a camada externa do cérebro, são responsáveis pelo processamento cognitivo.

As redes convolucionais combinam três ideias arquitetônicas para garantir algum grau de invariância de deslocamento e distorção: campos receptivos locais, pesos compartilhados e, às vezes, subamostragem espacial ou temporal (Yann Lecun, 2010).

A figura 13 representará de maneira simplificada o funcionamento de uma rede neural.

Uma imagem contendo interior, mesa, superior, sentado

Descrição gerada automaticamente

Figura 13: Rede neural simplificada. Fonte: Ferneda (2006)

De acordo com Moacir A. Ponti (2017), a principal aplicação para redes neurais objetiva o processamento de informações visuais, pois a convolução permite filtras imagens considerando sua estrutura bidimensional.

Um dos modelos para redes neurais convolucionais foi testado em 2012 pela ImageNet, no qual foram utilizadas 1,3 milhão de imagens, todas distribuídas em mais de 1000 classes diferentes. Cada imagem *RGB* foi pré-processada redimensionando a menor dimensão para 256, recortando a região central de 256x256, subtraindo a média por pixel (em todas as imagens) e usando 10 sub-culturas diferentes de tamanho 224x224 (cantos + centro com (fora) ) viradas horizontais (Zeiler & Fergus, 2012).

Segundo Ushizima (2017), as camadas convolucionais consistem de um conjunto de filtros que recebem como entrada um arranjo 3D, sendo chamado de volume, afirmando que cada filtro possui dimensão reduzida, porém pode se estender por toda a profundidade do volume de entrada. Caso uma imagem possua dimensão 5x5x3, automaticamente durante o processo de treinamento da rede, esses filtros são ajustados para que sejam ativados na presença de características relevantes identificadas no volume de entrada.

Na Figura 14 será mostrado, será mostrado uma ilustração de como serão formados os filtros convolucionais, formando mapas de caracteristicas.

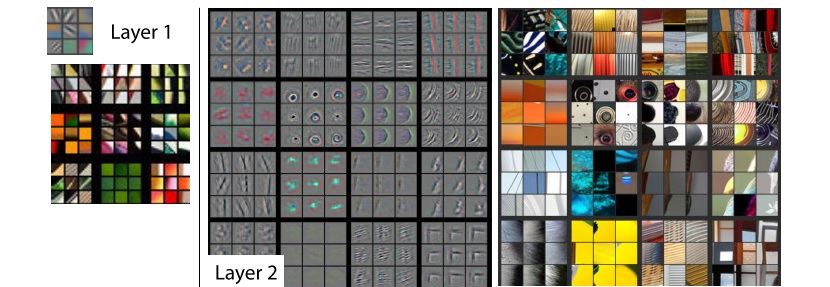


Figura 14: Etapas para utilização em redes neurais - Fonte: Zeiler & Fergus, 2012.

Cada filtro dá origem a uma estrutura conectada localmente que percorre toda a extensão do volume de entrada, os valores resultantes após a operação de convolução passam por uma função de ativação, uma das funções de ativação mais comuns é a função *ReLU (Rectified Linear Units)* (Ushizima, 2017). A equação 2.4 mostra a função *ReLU.*

|  |  |
| --- | --- |
|  | (2.4) |
|  |  |

A ideia de aprender a representação certa para os dados fornece uma perspetiva para o deep learning. Outra perspectiva sobre o deep learning é que a profundidade permite que o computador aprenda um programa de computador com várias etapas. Cada camada da representação pode ser considerada como o estado da memória do computador depois de executar outro conjunto de instruções em paralelo. Redes com maior profundidade podem executar mais instruções em sequência (Goodfellow, Ian, Yoshua Bengio, 2016).

Na Figura 15 será demonstrado como é realizado cada etapa de aprendizagem.

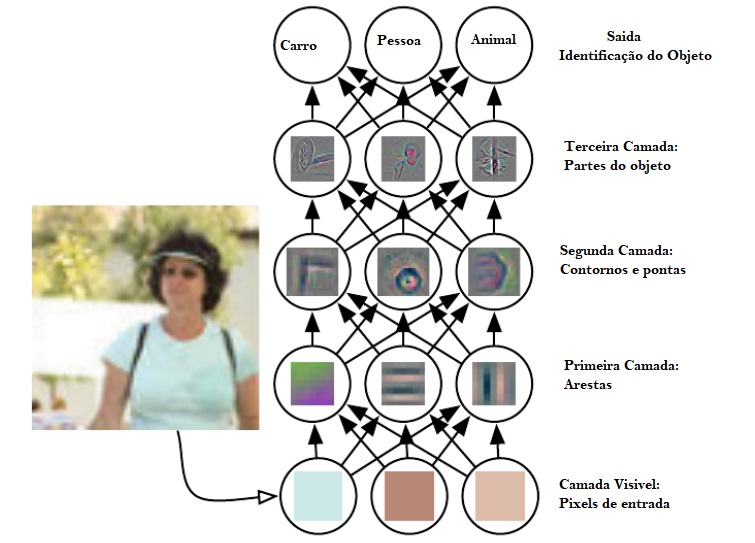


Figura 15: Ilustração das etapas de aprendizagem - Fonte: Goodfellow, Ian, Yoshua Bengio, 2016.

Segundo Silva (2015) deep learning trata-se de um conjunto de técnicas que utilizam redes neurais artificiais profundas, com muitas camadas intermediárias entre a camada de entrada e a de saída, afirmando também que o diferencial tecnológico dessa abordagem está nos excelentes resultados obtidos na resolução de problemas, resultados esses que superam até mesmo o desempenho dos melhores especialistas em determinadas áreas de conhecimento.

É importante considerar, também, o fato de que a estrutura hierárquica das camadas do córtex visual inspirou a criação de redes neurais convolucionais (Silva,2015).

A perspectiva da aprendizagem profunda é motivada por duas ideias principais. Uma ideia é que o cérebro fornece uma prova pelo exemplo de que o comportamento inteligente é possível, um caminho conceitualmente direto para a construção de inteligência é fazer engenharia reversa, um dos princípios computacionais por trás do cérebro e duplicar sua funcionalidade. Outra perspectiva é que seria profundamente interessante entender o cérebro e os princípios que fundamentam a inteligência humana, então os modelos de aprendizado de máquina que derramou luz sobre estas questões científicas básicas são úteis para além da sua capacidade de resolver aplicações de engenharia (Goodfellow, Ian, Yoshua Bengio, 2016).

Como base para as redes neurais a serem utilizadas no trabalho, foi abordada uma arquitetura de rede neural, que utiliza blocos de camadas executadas em paralelo, esses módulos são conhecidos como *Inception*, que consistem em combinações paralelas de camadas com filtros convolucionais de tamanho 1x1, 3x3 e 5x5. De Acordo com Ushizima (2017), a grande vantagem do módulo *Inception* é o uso da convolução com filtros 1x1, visando reduzir o número de características no bloco paralelo, antes de realizar a convolução com filtros maiores.

Uma imagem contendo texto

Descrição gerada automaticamente

Figura 16: Arquitetura do módulo *Inception* - Fonte: Ushizima (2017)

Baseado na arquitetura *Inception*, foi utilizado para esse trabalho duas arquiteturas de redes neurais, são elas as redes Xception (*Extreme Inception*) e a rede ResNet (*Residual Network*).

**Rede Neural Xception (*Extreme Inception*)**

Segundo Chollet (2014), a arquitetura *Xception* é uma rede composta por 36 camadas convolucionais formando a base de extração de características de rede. Comprova-se que a arquitetura é uma pilha linear de camadas de convolução separáveis em profundidade com conexões residuais. Dessa maneira se torna mais simples a configuração e modificação caso seja necessário.

A figura 17 mostrará a arquitetura de uma rede *Xception*, pode-se notar que todas as camadas de convolução separável usam um multiplicador de profundidade 1.

**Fluxo de Entrada Fluxo Médio Fluxo de Saída**

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Figura 17: Arquitetura *Xception*, onde os dados passam pelo fluxo de entrada, posteriormente pelo fluxo médio e por fim passa pelo fluxo de saída. Fonte: Chollet (2014)

A arquitetura *Xception* utiliza 3 fluxos como método de aprendizado, de acordo com a figura 17, foi utilizado como entrada uma imagem de tamanho 299x299x3, 3 sendo o número de canais de cada pixel da imagem *RGB*. O fluxo de entrada é fase que se geram os mapas de características, ou seja, a saída do fluxo de entrada são os mapas de características, fazendo convolução com filtro 1x1, utilizando *stride* de 2x2. *Stride* é o número de posições em que serão movimentados pelo filtro convolucional. Como entrada para o fluxo médio, são os mapas de características, na figura 17 são apresentadas imagens com tamanho 19x19, separadas em 728 mapas de características, formando dessa maneira nós empilhados 19x19x728.

**Rede Neural ResNet (*Residual Network*)**

De acordo com Ushizima (2017), uma rede *ResNet* é composta por 152 camadas e formada por blocos residuais, afirmando que o propósito dos blocos residuais são os de que, uma determinada entrada , passará por uma série de operações de convolução.

Considerando como um mapeamento subjacente a ser ajustado por camadas empilhadas, sendo x o número de entradas para primeira camada, então é equivalente a hipótese de que é possível se aproximar assintoticamente as funções residuais , partindo da ideia que a entrada e a saída são das mesmas dimensões. Assim não se deve esperar camadas empilhadas para aproximar , utiliza-se uma função residual ,assim a função original é (He & Sun, 2015). O resultado da operação é adicionado a entrada original de , como é ilustrado na figura 18.

Na arquitetura ResNet são utilizados resultados de camadas anteriores como uma segunda entrada para uma nova camada, esse segundo valor, pode ser chamado como bloco residual (He & Sun, 2015). O bloco residual será somado ao valor de saída de uma determinada camada da rede, essa afirmação pode ser comparada ao processo de entrada de valores de um neurônio. Ao fim da função de ativação mostrada na seção sobre neurônios, é somado um valor a saída, esse valor é chamado de bias A soma de um valor do bloco residual pode ser assimilada a soma do valor *bias* no valor de saída de um neurônio artificial (He, K., & Sun, J. 2015).

A figura 18 mostra como ocorre o aprendizado residual, *weight layer* é peso da camada e x o valor identidade, ou o valor do bloco residual.

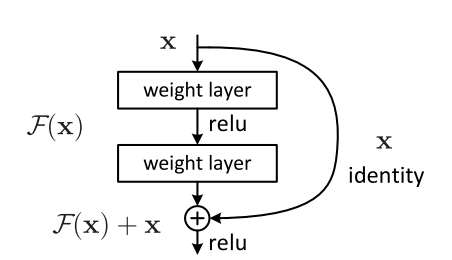


Figura 18: Demonstração de como é entrado cada valor nas camadas da rede ResNet– Fonte: He & Sun (2015)

## Métricas de Avaliação

Métricas de avaliação serão utilizadas para avaliar caracteristicas de determinados objetos, por exemplo, ao comparar a precisão dos algoritmos de reconhecimento de objetos, é normalmente aceitável pré-processar as imagens de entrada de forma ligeiramente diferente para cada algoritmo, com base nos tipos de requisitos de entrada que ele possui. A modelagem gerativa é diferente porque as mudanças no pré-processamento, mesmo as muito pequenas e sutis, são completamente inaceitáveis.

Qualquer alteração nos dados de entrada altera a distribuição a ser capturada e altera fundamentalmente a tarefa (Goodfellow, Ian, Yoshua Bengio, 2016).

Na área de recuperação de informação, precisão é um número de elementos relevantes recuperados, dividido pelo número total de elementos recuperados, revocação é definida pelo número de elementos relevantes recuperados pelo número total de elementos relevantes existentes (Matos, 2009).

Os valores são representados por uma matriz de confusão, a figura 19 mostra a determinada matriz de confusão.

Uma imagem contendo captura de tela

Descrição gerada automaticamente

Figura 19: Matriz de confusão - Fonte: Fawcett (2006)

Sendo a classe de valores verdadeiros positivos, valores falsos positivos, valores falsos negativos e valores verdadeiros, porém negativos.

Precisão e revocação na área de inteligência artificial utiliza a matriz de confusão apresentada na figura 19, que obtém a medida efetiva do modelo de classificação, ao mostrar o número de classificações corretas e as classificações preditas para cada classe em uma determinado conjunto de exemplos (Matos, 2009).

|  |  |
| --- | --- |
|  | (2.5) |
|  | (2.6) |

*Mean Absolute Error* (MAE) ou Erro Médio Absoluto produz uma função que prediz o valor mediano de y para cada x, desde que tal função possa ser descrita pela família de funções sobre as quais otimizamos. Esta função de custo é comumente chamada erro absoluto médio. Infelizmente, o erro quadrático médio e o erro absoluto médio geralmente levam a resultados quando usado com otimização baseada em gradiente. Algumas unidades de saída que saturam produzem gradientes muito pequenos quando combinadas com essas funções de custo (Goodfellow, Ian, Yoshua Bengio, 2016).

|  |  |
| --- | --- |
|  | (2.7) |

*F-Score* ou Medida-F é quando a precisão e a revocação tem o mesmo peso, uma medida da acurácia do teste.

|  |  |
| --- | --- |
|  | (2.8) |

*Log Loss* é a função que pune previsões incorretas, como por exemplo, prever que uma determinada classe tem 90% de probabilidade, quando na verdade está incorreto.

|  |  |
| --- | --- |
|  | (2.9) |

Acurácia ou precisão geral é o número de acertos positivos divido pelo número total de exemplos.

|  |  |
| --- | --- |
|  |  |
|  | (2.9.1) |
|  |  |

****VALIDAÇÃO CRUZADA****

Como a saída das redes neurais são valores em probabilidade, pode-se utilizar um método de verificação da veridicidade dos dados obtidos, a partir de uma validação cruzada.

Uma das técnicas mais utilizadas para verificar a estabilidade de uma estrutura é a validação cruzada. Esta técnica permite examinar se a estrutura identificada se repete quando investigada em uma segunda amostra. A sua utilização é importante tanto para soluções fatoriais exploratórias como para soluções fatoriais confirmatórias (A. Jacob, 2004).

## Trabalhos relacionados a detecção de fogo

Nessa seção será mostrado alguns trabalhos relacionados como detecção de fogo em vídeos utilizando métodos para considerar as mesmas como fogo. (B. Ugur Toreyin, Yigithan Dedeoglu, Ugur Gudukbay (2006), Celik, Demirel, Ozkaramanli, & Uyguroglu (2007), Ã, Cheong, & Nam (2009), Liu & Ahuja (2009), (Ko, Cheong, & Nam, 2010)).

No trabalho realizado por B. Ugur Toreyin, Yigithan Dedeoglu, Ugur Gudukbay (2006), foi idealizado um novo método para detecção de chamas em tempo real, captando os vídeos gerados por uma câmera de monitoramento, no trabalho já sabendo sobre informações comuns como movimento e destaque das cores, as chamas foram detectadas pela análise do vídeo no domínio de uma *wavelet*. As detecções de chama são realizadas de acordo com um cálculo que transformam *wavelets* em regiões onde supostamente podem haver fogo.

No trabalho de Celik, Demirel, Ozkaramanli, & Uyguroglu (2007), foi proposto um detector de incêndio em tempo real que combina informações de objeto em primeiro plano com estatísticas de cores em pixels do fogo.

Modelo de fundo adaptativo simples da cena é gerado usando três distribuições gaussianas, na qual cada distribuição corresponde à estatística de pixel no respectivo canal de cor. As informações de primeiro plano foram extraídas usando algoritmo de subtração de fundo adaptável e, em seguida, verificadas pelo modelo de cor de fogo estatístico para determinar se o objeto de primeiro plano detectado é um candidato a incêndio ou não. Um modelo genérico de cores de fogo é construído pela análise estatística das imagens de amostra contendo pixels de fogo. O artigo também separa em contribuições de acordo com a aplicação do método utilizado.

Já no trabalho realizado por Ã, Cheong, & Nam (2009), foi sugerido um novo método de detecção de incêndio baseado em sensores de visão para um sistema de monitoramento de incêndio de alerta antecipado. Primeiro, as regiões de incêndio candidatas são detectadas utilizando versões modificadas de métodos relacionados anteriores, como a detecção de regiões em movimento e pixels cor de fogo. Em seguida, como as regiões de fogo geralmente têm um contraste de luminância mais alto do que as regiões vizinhas, um mapa de luminância é feito e usado para remover os pixels que não são de fogo. Posteriormente, um modelo de incêndio temporal com coeficientes de wavelets é criado e aplicado a um classificador de máquinas de vetores de suporte de duas classes (SVM) com um kernel de função de base radial (RBF). O classificador SVM é então usado para a verificação final do pixel de fogo.

No trabalho de Liu & Ahuja (2009), foram propostos modelos espectrais, espaciais e temporais de regiões de fogo em sequências de imagens visuais. O modelo espectral foi representado em termos da densidade de probabilidade de cor dos pixels de fogo. O modelo espacial captura a estrutura espacial dentro de uma região de fogo. A forma de uma região de fogo é representada em termos do conteúdo de frequência espacial do contorno da região usando seus coeficientes de Fourier. As mudanças temporais nestes coeficientes são usadas como as assinaturas temporais da região do fogo. Especificamente, um modelo autoregressivo da série de coeficientes de Fourier é usado.

No trabalho realizado por (Ko et al., 2010), foi proposto um novo método de detecção precoce de incêndio baseado em visão para aplicação no mundo real. Primeiro, as regiões de fogo do candidato são detectadas usando um modelo de fundo e cor de fogo. Modelos probabilísticos do fogo são então gerados com base no fato de que os valores de pixels de fogo em quadros consecutivos mudam constantemente. Esses modelos são então aplicados às redes bayesianas. Este artigo usa redes bayesianas hierárquicas que contêm nós intermediários. Quatro funções de densidade de probabilidade para evidência em cada nó são usadas. As funções de densidade de probabilidade para cada nó são modeladas usando a assimetria da cor vermelha e três altas freqüências obtidas de uma transformada *wavelet*.

# Metodologia

O escopo do trabalho é a detecção de fogo em vídeos, porém, para encontrar píxeis que tenham porcentagem significativa de chances de ser classificado como fogo, é necessário passar por uma cadeia de processos. Neste sentido, o presente capítulo tem como objetivo apresentar os métodos adotados para que fosse possível alcançar os objetivos. A figura 20 ilustra as etapas necessárias para a construção de um sistema de detecção de fogo e fumaça.

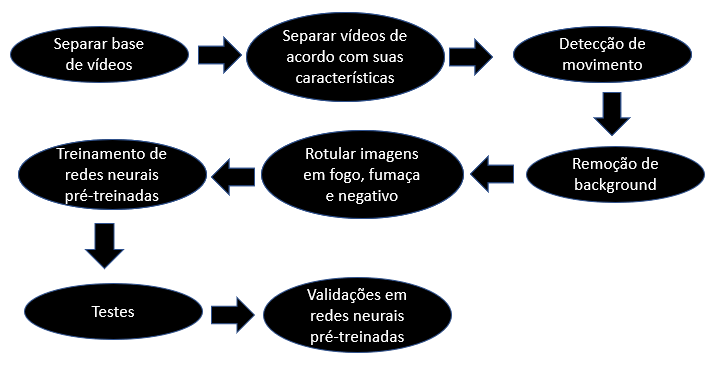


Figura 20: Metodologia adotada para realização do trabalho

Primeiramente é necessária uma base de vídeos, onde a câmera não se mova, mas sim os objetos dentro do vídeo, sendo assim possível, se detectar movimento. Posteriormente é necessário um método que consiga, através dos valores dos píxeis em determinado tempo t, detectar que há movimento dentro do vídeo.

Segundo Chen et al. (2010), o primeiro passo é alcançado através da aplicação de um modelo de mistura gaussiana GMM (*Gaussian Mixture Model)* para analisar o espaço de cor *RGB* dos pixels da imagem, utilizando esse método é possivel se distinguir dentro de uma imagem o que é fundo e o que é frente, tornando mais simples a detecção de movimento.

Pela eficiencia do metodo GMM descrito por (Steffens, 2015), foi então adotado esse método para realização de reconhecimento de *backgound*.

A partir da detecção de movimento é possivel retirar imagens dos vídeos, somente das partes que interessam, ROI (*Regions Of Interests*), sendo assim, retirada todo o fundo dos videos e mantendo nas imagens, somente partes onde se encontrou movimento. Após a extração das imagens, removendo locais das imagens que foram caracterizados como estático, para serem utilizados como entrada para o treinamento das redes convolucionais. O processo de monstagem dessa base será descrito da subção a seguir.

## Base de Imagens a partir de vídeos

Para início do trabalho foi levantada uma base de vídeos disponibilizadas pelo trabalho realizado por Foggia, P., Saggese, A., & Vento, M. (2015) para serem trabalhados com os mesmos, no momento de testes, para a detecção de movimento, remoção de *background* e separação de fundos.

Nessa base de vídeos, foram escolhidos vídeos onde o cenário apresentado era o de uma câmera estática, onde objetos se movimentavam, havia chamas em determinados vídeos, haviam fumaça em outros vídeos, haviam movimentações nas imagens dos vídeos, porém a posição da câmera permanecia estática. A princípio haviam 33 vídeos na base, após selecionar os vídeos que se encaixavam melhor nas características necessárias, foram mantidos apenas 24 vídeos. A Figura 20 demonstra 4 *frames* dos vídeos selecionados.

Figura 21: *Frames* dos vídeos selecionados para o trabalho

Após a seleção de 29 vídeos 17 vídeos continham fogo e 12 vídeos continham fumaça. O próximo passo foi o de detectar o movimento dentro dos vídeos, para isso foi utilizado um algoritmo na linguagem python3, com a biblioteca OpenCV, essa biblioteca permite que seja possível o processamento, manipulação e analise de imagens, com ela é possível ter uma visão computacional das imagens.

O código basicamente abre um clipe, e detecta se dentro desse clipe há movimento, utilizando os pixels anteriores e as comparações de mudanças temporais aplicando o filtro de desfoque gaussiano (*Gaussian Blur*) com máscara de tamanho 7x7 para destacar regiões de cores distintas, após isso, aplicou-se uma máscara baseada em operações de dilatação da morfologia matemática.

A figura 21 indica como ficaram as imagens após a detecção de movimento:

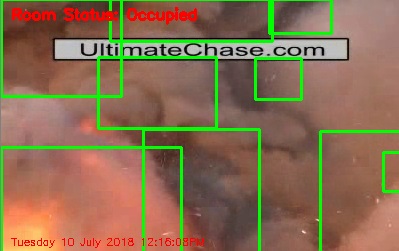
 

Figura 22: Imagens dos *frame*s onde foi detectado movimento

## Detectando regiões de interesse

O terceiro processo do trabalho foi a remoção dos fundos dos vídeos, após detectar o movimento, ficaria mais simples de se identificar o fundo das imagens e conseguir separar a frente da imagem do fundo, foi mantido como *foreground* todos os objetos que se movimentavam e foi definido como *background* os objetos que permaneciam-se estáticos.

A técnica utilizada para a remoção de *background* foi a de obter uma máscara como *threshold*, mostrando onde há movimento. Para cada *frame* há uma máscara *threshold* e um *frame* original sem nenhuma modificação, utilizando a máscara dentro do *frame* original foi selecionado apenas a área do *frame* onde se houve movimento, após isso foi retirado do quadro tudo que não estava dentro da máscara. As imagens foram gravadas a cada 10 iterações, o que seriam aproximadamente 3 *frames* por segundo, considerando 30fps (*frames* por segundo), o código na figura 23 demonstra como foi realizado esse processo.

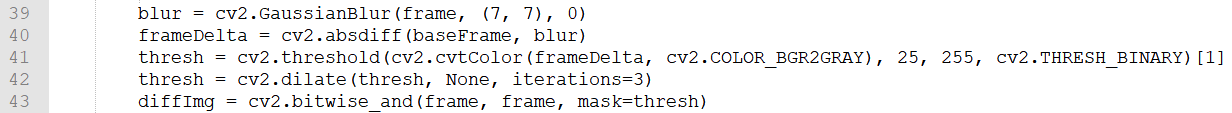


Figura 23: Código utilizado para gravar imagens já processadas

A Figura 24 mostra 4 *frames* dos vídeos já sem fundo, foi deixado como preto a parte em que foi identificado como estático.

Figura 24: Imagens já processadas e apenas com movimento detectado

Após a remoção dos fundos dos *frames*, foi mantido em pastas diferentes imagens com fogo, com fumaça e sem ambas. Posteriormente essas imagens sem fundo, ou seja, onde não foi caracterizado como movimento, foram aplicados os valores dos píxeis como 0, deixando a área onde não há movimento, preto. Essas imagens serão utilizadas para serem aplicadas dentro de uma base, para serem consumidas por uma rede neural convolucional para então haver a extração de características de acordo com os filtros convolucionais e a aprendizagem profunda.

## Aplicação das imagens em redes neurais

Posteriormente a separação das imagens que possuem apenas o *foreground*, essas mesmas, foram aplicadas a uma base, em redes neurais, para serem passadas por todas as camadas e então ser avaliado percentualmente a probabilidade dessa imagem possuir fogo. Para aplicar as redes neurais, a base de imagens foi dividida em base de teste e validação, isto possibilitou a realização do treinamento e os testes das redes neurais estudadas. Foram separadas em pequenas porcentagens as imagens para teste e validação nas redes neurais, separando as imagens em pastas, uma pasta somente com imagens para validação e outra somente para teste, fazendo desta forma pode-se evitar com que a rede neural fique viciada, ou seja, apenas reconhecendo imagens que possam haver fogo. Dessa maneira, foi adotada a validação cruzada, dividida em duas repartições, uma base de teste e uma de validação. A validação cruzada visa que os resultados obtidos pelas redes neurais, independente da sua arquitetura, sejam avaliados como uma prova real, para haver certeza de que a rede neural, não esteja viciada.

As arquiteturas de redes neurais utilizadas no trabalho foram as redes *Xception* e *ResNet*, nessas redes foram realizados treinos com camadas finais totalmente conectadas com 1024 neurônios em cada camada. Para o processo de treinamento das redes neurais, foi utilizado um *patche* com 110 mil imagens de fogo e fumaça. Na base de treinamento foram realizadas 10 iterações, isolando a base de validação, para posteriormente haver um processo com a rede inteira. Esse processo foi realizado com as duas redes neurais.

# Desenvolvimento

De acordo com a metodologia apresentada na seção 3, a realização do desenvolvimento do trabalho ocorreu da seguinte maneira. Primeiramente foi recolhido 33 vídeos com características semelhantes, sendo elas, vídeos onde a câmera não se movimenta mantendo uma posição somente para a câmera, as imagens que são reproduzidas nos vídeos foram utilizadas como parâmetro para separação dos vídeos. Esses vídeos foram disponibilizados no trabalho realizado por Foggia, P., Saggese, A., & Vento, M. (2015).

Para cada vídeo foi selecionado uma característica, sendo essas características haver fogo, fumaça e não haver nem fogo e nem fumaça.

Após fazer o reconhecimento dessas características foram descartados 4 vídeos dos 33 separados, pois não atendiam as necessidades requeridas.

Com uma nova base havendo 29 vídeos, foi realizada então a separação dos vídeos com suas características, assim então separados os vídeos em pastas diferentes.

Após a separação dos vídeos em pastas especificas de acordo com as características dos mesmos, foi montado um algoritmo para realizar a detecção de objetos se movimentando.

O algoritmo foi desenvolvido na linguagem Python3 por maiores afinidades com a linguagem e pela maior facilidade em processamento de imagens tendo como biblioteca base para métodos de detecção de movimento OpenCV, essa biblioteca que possibilita ao desenvolvedor uma visão computacional dos vídeos, sendo possível manipular e analisar as imagens dos vídeos.

Após a escolha da linguagem que será desenvolvido e a biblioteca que auxiliara com os métodos, foi desenvolvido um algoritmo que carrega um vídeo e salva imagens onde foi encontrado movimento. Para a detecção de movimento foram utilizados métodos já prontos trazidos pela biblioteca OpenCV. O primeiro método utilizado foi o *Gaussian Bluring*, esse método foi eficaz para remoção de ruídos em imagens.

Uma imagem contendo céu, ao ar livre, árvore, neve

Descrição gerada automaticamente

Figura 25: Mostrando o resultado obtido após a utilização do método GB (*Gaussian Blur*).

Após ser utilizado esse método foi calculado a diferença absoluta do *frame* gaussiano e do *frame* original, o resultado servirá como parâmetro para cálculo de um valor para limiar (*Threashold*), no próximo método.

Uma imagem contendo televisão, interior

Descrição gerada automaticamente

Figura 26: Resultado após a utilização do método para diferença absoluta

Posteriormente é necessário a utilização do método de *threshold*, essa função aplicará o limite dos níveis fixos de uma matriz de vários canais, essa função foi utilizada para resultar uma imagem de dois níveis, ou seja, uma imagem binária. Porém, primeiro foi necessário transformar a soma dos dois *frame*s o primeiro *frame* original sem nenhuma modificação e o *frame* com a diferença absoluta, isso foi necessário para transformar os dois *frame*s em tons de cinza.



Figura 27: Imagem utilizada como máscara para detecção de movimento

Após gerar imagens com *frames* binário, foi escolhido método, pois pode-se notar que onde existem píxeis brancos na imagem foi onde se encontrou movimento partindo da comparação entre *frame*s passados. Pode-se notar que as imagens foram dilatadas utilizando o método de dilatação da biblioteca.

Após realizar a dilatação dos píxeis em branco, foi necessário se fazer uma comparação entre os píxeis onde se tem movimento em preto e branco, com os píxeis do vídeo colorido. Após essa comparação foi possível remover o *background* dos vídeos e manter apenas onde se encontrou movimentação.

Uma imagem contendo bolo, mesa, animal, interior

Descrição gerada automaticamente

Figura 28: Imagem final após passar pelo processo de detecção de movimento e remoção de *background*

Após ter separado os vídeos de acordo com a seção 3, a próxima parte do trabalho foi a de utilização das redes neurais *Xception* e *ResNet* para treinamento, testes e validação dos resultados obtidos com as imagens que foram utilizadas como entrada.

Após o treinamento com as camadas totalmente conectadas, foi realizado então 10 iterações para um pré-treinamento na rede *Xception*, os gráficos a seguir mostram os valores da acurácia, do F-Score e o *Log Los*s do aprendizado da rede de acordo com as iterações, os valores estão em porcentagem conforme as métricas de avaliação relatadas na seção 2.5.

Figura 29: Acurácia em relação a iterações na base de treino utilizando os *frame*s de vídeos

Figura 30: F-Score em relação a iterações na base de treino utilizando os *frame*s de vídeos

Figura 31: Log *Los*s em relação a iterações na base de treino utilizando os *frame*s de vídeos

Figura 32: Erro médio absoluto em relação a iterações na base de treino utilizando os *frame*s de vídeos

A tabela a seguir traz os valores da Acurácia, *F-Score,* *Loss* e *Mean Absolute Error* recolhidos pelo pré-treinamento realizado pela rede neural *Xception*.

**Tabela 1 – Valores recolhidos pela rede *Xception* na base de teste**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Número de iterações | Acurácia | *F-Score* | *Loss* | *Mean Absolute Error* |
| 1 | 0,510229 | 0,501209 | 2,011356 | 0,360095 |
| 2 | 0,751805 | 0,749165 | 0,567813 | 0,222631 |
| 3 | 0,829723 | 0,819756 | 0,420981 | 0,170846 |
| 4 | 0,864922 | 0,864456 | 0,331502 | 0,135122 |
| 5 | 0,895608 | 0,893882 | 0,272562 | 0,111941 |
| 6 | 0,921179 | 0,920135 | 0,224952 | 0,090589 |
| 7 | 0,919073 | 0,919543 | 0,214249 | 0,088095 |
| 8 | 0,933514 | 0,933019 | 0,198012 | 0,078772 |
| 9 | 0,930806 | 0,92909 | 0,186854 | 0,075758 |
| 10 | 0,929904 | 0,93 | 0,198082 | 0,079781 |

De acordo com os valores apresentados nos gráficos e na tabela, o pré-treinamento realizado na rede neural *Xception*, na base de treinamento, foi de extrema qualidade, com valores de acurácia chegando a um valor de 93,35%, valores significativos para o propósito do trabalho. Outro ponto interessante, foram os valores da métrica F-Score, alcançando uma marca de 93,30%, ou seja, essas iterações chegaram à conclusão de que as imagens utilizadas na entrada para serem processadas pelas redes neurais, possuem 93,30% de chances de serem fogo, tomando em consideração que os valores apresentados pela métrica *Loss*, são valores baixos, principalmente valores após a 6ª iteração, chegando a marca de somente 19% de chance deste processo de aprendizado estar errado.

Após o pré-treinamento nas redes, foram então realizadas 40 iterações na mesma rede neural *Xception*, desta vez com a rede neural completa, utilizando a base de validação. Os gráficos a seguir mostram o crescimento da acurácia e do *F-Score*, *Log Loss* e *Mean Absolute Error* após o pré-treinamento da rede.

Figura 33: Acurácia após ser aplicado *frame*s na base de validação

Figura 34: F-Score após ser aplicado os *frame*s na base de validação

Figura 35: *Log Loss* após ser utilizado os *frame*s na base de validação

Figura 36: Erro Médio Absoluto após ser aplicado os *frame*s na base de validação

**Tabela 2: Melhores valores recolhidos pela rede neural *Xception*, na base de validação**

|  |  |  |  |
| --- | --- | --- | --- |
| Acurácia | *F-Score* | *Loss* | *Mean Absolute Error* |
| 0,98952 | 0,98852 | 0,046085 | 0,008633 |

Os valores da tabela 2, são os melhores valores recolhidos da base de validação da rede *Xception*. De acordo com a rede as imagens que foram separadas para serem utilizadas como entrada, possuem de 97% a quase 99% de serem imagens contendo chamas. Os resultados apresentados na Acurácia e no F-Score são os números que apoiam a tal informação, mas não somente eles, os valores para erro médio absoluto e o *Loss*, também estão mostrando a possibilidade entre 97% de chances de tais imagens separadas possuírem imagens de fogo, *Loss* indicando 3% de chances do processo estar equivocado e o erro absoluto mínimo apenas 1%, ou seja, números indicam que os vídeos possuem uma grande chance de possui fogo.

Após os resultados apresentados pela rede *Xception*, foi então a vez de utilizar a rede neural *ResNet* para realização de um novo processo, utilizando as mesmas métricas de avaliação, porém em uma nova arquitetura de rede convolucional. Na rede *ResNet* foram realizadas 4 iterações para treinamento das camadas totalmente conectadas, carregando os pesos da rede que já havia sido pré-treinada com um patch de 110 mil imagens de fogo e fumaça, com os mesmos processos da rede *Xception*, um pré-treinamento com a base de treinamento e posteriormente utilizar a base de validação. Os gráficos e a tabela a seguir mostram os resultados recolhidos pela *ResNet*.

Figura 37: Acurácia em relação a iterações na base de validação utilizando os *frame*s de vídeos na rede ResNet

Figura 38: F-Score em relação a iterações na base de validação utilizando os *frame*s de vídeos na rede ResNet

Figura 39: *Log Loss* em relação a iterações na base de validação utilizando os *frame*s de vídeos na rede *ResNet.*

Figura 40: Erro Médio Absoluto em relação a iterações na base de validação utilizando os *frame*s de vídeos na rede *ResNet.*

A rede apresentou um resultado para Acurácia e F-Score entre 97 a 99% de chances das imagens recolhidas pelos vídeos possuem chamas, afirmando mais ainda que esse processo esteja correto são as métricas *Loss* e *Mean Absolute Error*, mostrando respectivamente valores entre 3 a 4% e de 1 a 2% em uma média de erros absolutos. Baseado nos resultados apresentados pela rede *Xception* e a rede *ResNet*, pode-se afirmar que os vídeos escolhidos possuem imagens que contem chamas. Tais valores representados nos gráficos foram retirados dos resultados indicados pela base de treinamento da rede, a tabela 3 mostra os valores da base de validação.

**Tabela 3: Melhores valores recolhidos pela rede *ResNet***

|  |  |  |  |
| --- | --- | --- | --- |
| Acurácia | F-Score | *Loss* | *Mean Absolute Error* |
| 0,989771 | 0,989919 | 0,03515 | 0,012143 |

Na base de validação as métricas de Acurácia e F-Score chegaram a valores de 100%, ou seja, a base reconheceu todas as imagens passadas durante a iteração, as métricas *Loss* e erro médio absoluto ficaram com valores significativamente baixos. Com tais valores, é possível afirmar, que os vídeos utilizados na base, possuem imagens de fogo. Fazendo uma comparação entre bases de validação nas redes, a base com o valor mais significativo, foi a base de validação da rede *ResNet*, chegando a marca de 100% de acerto.

# Conclusão

Após o desenvolvimento ter chegado a valores significativos para o trabalho, foi concluído que as redes neurais utilizadas, *ResNet* e *Xception*, resultaram em valores positivos para o trabalho, a rede *ResNet* conseguindo valores significativos, acertando 100% em uma iteração, para a acurácia e *F-Score*, como demonstra a tabela 3. Concluindo que as redes neurais foram utilizadas de maneira correta para chegar a tal resultado, deixando uma base de teste e uma base para validação, tornando mais eficiente o trabalho de reconhecimento de chamas, partindo também do pressuposto que foram utilizadas 110 mil imagens para treinamento, essas imagens possuíam chamas e fumaça. Os problemas encontrados durante o trabalho, foram os de métodos para realização da detecção de movimento, alguns trabalhos relacionados não possuíam uma didática muito boa, tornando mais complexo o entendimento das funções para reconhecimento de fundo, em alguns momentos foram utilizados métodos que não possuíam filtros gaussianos, mas que utilizavam valores limiares, deixando a imagem preto e branco (imagem binária), porém foi realizado um processo com imagens com canais de cor *RGB* e assim foi concluído com sucesso retornando imagens coloridas, detectando onde havia movimento, a detecção do movimento foi realizada a partir de uma máscara de valores limiares aplicado nas imagens com valores *RGB*. Problemas relacionados a material, uma boa parte do material encontrado para formular a fundamentação do trabalho era em outra linguagem de domínio, tornando um pouco mais difícil o entendimento sobre o assunto. Devido a ser um tema relativamente novo, como arquiteturas de redes novas, encontrar material conceituando as redes também foi difícil e por motivos de ser um processamento enorme de imagens, utilizando um grande número de parâmetros, foi necessária uma máquina que fosse capaz de processar todos os dados juntamente a rede neural. Para trabalhos futuros, pode-se utilizar novas arquiteturas de redes neurais, não se limitando em duas redes, também pode ser utilizado mais vídeos na base, assim sendo possível utilizar mais *frame*s para testes, treinamento e validação, também podem ser utilizados novos métodos de detecção de movimento, ou métodos para detecção de chamas utilizando valores dos píxeis. Pode-se concluir que o trabalho chegou ao seu propósito, o de detectar fogo em vídeos utilizando redes neurais convolucionais, extraindo *frame*s de vídeos, de 3 em 3 segundos partindo da ideia de ser passado 30 *frame*s por segundo. Conclui-se que o propósito foi apresentado de maneira positiva, baseando-se nos valores gerados pelas redes neurais, *Xception* e *ResNet*.

# Referencias

Ã, B. C. K., Cheong, K., & Nam, J. (2009). *Fire detection based on vision sensor and support vector machines*, *44*, 322–329.

A. Jacob Laros, E. Katia Puente-Palacios (2004). Validação cruzada de uma escala de clima organizacional.

Barreto, J. M. (2002). Introdução a redes neurais artificiais.

Carlin, N., Szanto, E. M., Seale, W. A., Jorge, F. O., Souza, F. A., & Gasques, I. H. B. L. R. (2004). Processamento de imagens : Conceitos básicos relacionados com o uso de um computador óptico fênomeno da difração, 241–245.

Castro, C. F. De, Serra, G., Parola, J., Reis, J., Lourenço, L., Castro, C. F. De, … Abrantes, J. B. (2003). Combate a incêndios florestais (Vol. XIII).

Celik, T., Demirel, H., Ozkaramanli, H., & Uyguroglu, M. (2007). *Fire detection using statistical color model in video sequences*, *18*, 176–185.

Chen, J., He, Y., & Wang, J. (2010). *Multi-feature fusion based fast video flame detection. Building and Environment*, *45*(5), 1113–1122.

Chollet, F. (2014). *Xception: Deep Learning with Depthwise Separable Convolutions.*

Corrêa, C., & Braga, G. C. (2007). Mapeamento de Incêndios em Edificações : um estudo de caso na cidade do Recife, *2*(3), 15–34.

Dos, D., & Gomes, S. (2010). Inteligência Artificial : Conceitos e Aplicações, 234–246.

Eustáquio, J., Queiroz, R. De, & Gomes, H. M. (2001). Introdução ao Processamento Digital de Imagens Considerações Iniciais, 1–31.

Facon, J. (2011). A Morfologia Matemática e suas Aplicações em Processamento de Imagens, 61–128.

Fawcett, T. (2006). *An introduction to ROC analysis*, *27*, 861–874.

Ferauche, T., & Almeida, M. A. De. (2011). Utilização de Técnicas de Inteligência Artificial na tarefa de Classificação Jurisprudencial do Tribunal Regional do Trabalho 2a . Região São Paulo, 59–70.

Ferneda, E. (2006). Redes neurais e sua aplicação em sistemas de recuperação de informação, 25–30.

Fiedler, M. B. de M. N. C. (2003). INCÊNDIOS FLORESTAIS NO PARQUE NACIONAL DA SERRA DA CANASTRA: DESAFIOS PARA A CONSERVAÇÃO DA BIODIVERSIDADE, 157–168.

Goodfellow, Ian, Yoshua Bengio, A. C. (2016). *Deep Learning*. MIT.

He, K., & Sun, J. (2015). *Deep Residual Learning for Image Recognition*, 1–9.

Jr, R. C. (2015). A Utilização de Filtros Gaussianos na Análise de Imagens Digitais, *3*(XXXV), 1–7.

Ko, B., Cheong, K., & Nam, J. (2010). *Early fire detection algorithm based on irregular patterns of flames and hierarchical Bayesian Networks. Fire Safety Journal*, *45*(4), 262–270.

Liu, C., & Ahuja, N. (2009). *Vision Based Fire Detection*, (1).

Luiz, A., Tebaldi, C., Fiedler, N. C., Juvanhol, R. S., & Dias, H. M. (2013). Ações de Prevenção e Combate aos Incêndios Florestais nas Unidades de Conservação Estaduais do Espírito Santo. *Actions of Prevention and Combat to Forest Fires in Conservation Units from Espirito Santo state , Brazil,* *20*(4), 538–549.

Matos, P. F. (2009). Universidade de São Paulo ( USP ) Universidade Federal de São Carlos ( UFSCar ) Universidade Metodista de Piracicaba ( Unimep ) Projeto - Um Ambiente para Análise de Dados.

Moacir A. Ponti, G. B. P. da C. (2017). Como funciona o Deep Learning.

Rauber, T. W. (2010). Redes Neurais Artificiais.

Scuri, A. E. (2002). Fundamentos da Imagem Digital.

Seito, A. I., Gill, A. A., Pannoni, F. D., Ono, R., Bento, S., Carlo, U. Del, & Pignatta, V. (2008). A Segurança Sobre Incêndio no Brasil. São Paulo: Projeto Editora.

Silva, L. A. D. e. (2015). Utilização de deep learning em ações de controle.

Sonaglio, S., & Moecke, E. M. (2008). Detecção de movimento para sistema automático de vigilância por vídeo, 35–43.

SP, C. T. P. D. N. D. S. E. N. E. DE. (2010). TEORIA DO FOGO Conceito de Fogo Elementos que compõem o fogo Sólidos Líquidos, pp. 231–242.

Steffens, C. R. (2015). Um Sistema de Detecção de Fogo Baseado em Vídeo.

Tamiozzo, F., Torres, P., Paulina, M., Roque, B., Lima, G. S., & Martins, S. V. (2017). Mapeamento do Risco de Incêndios Florestais Utilizando Técnicas de Geoprocessamento *Mapping of Forest Fires Risk Using Geoprocessing Techniques*, *8087*.

Técnico, R., Máquina, A. De, & Falciforme, D. A. (2009). Universidade de São Paulo ( USP ) Universidade Federal de São Carlos ( UFSCar ) Universidade Metodista de Piracicaba ( Unimep ) Projeto - Um Ambiente para Análise de Dados.

Ushizima, D. M. (2017). Redes Neurais Convolucionais com Tensorflow: Teoria e Prática, 382–406.

Yann Lecun, Y. B. (2010). *Convolutional Networks for Images, Speech, and Time-Series.*

Zeiler, M. D., & Fergus, R. (2012). *Visualizing and Understanding Convolutional Networks.*

Zivkovic, Z. (2004). *Improved Adaptive Gaussian Mixture Model for Background Subtraction*, (2).