

UNIVERSIDADE FEDERAL DE SÃO CARLOS - CAMPUS SOROCABA

CCGT - CIÊNCIA DA COMPUTAÇÃO

Processamento de Imagens

Prof. Dr. Jurandy Almeida

## **PROJETO DE PROCESSAMENTO DE IMAGENS**

Marco do Projeto

Felipe Bonadia de Oliveira Bravo - 813908

João Vitor Naves Mesa - 814149

Mateus Tsuyoshi Matsuo Hashimoto - 813500

## SUMÁRIO

<b>1. INTRODUÇÃO.....</b>	<b>3</b>
<b>2. DEFINIÇÃO DO PROBLEMA.....</b>	<b>3</b>
<b>3. ABORDAGEM TÉCNICA.....</b>	<b>3</b>
<b>4. RESULTADOS PRELIMINARES.....</b>	<b>4</b>
<b>5. ANEXOS.....</b>	<b>6</b>
<b>6. REFERÊNCIAS.....</b>	<b>6</b>

## 1. INTRODUÇÃO

A detecção precoce de fumaça e fogo é crucial para evitar grandes incêndios em ambientes industriais e residenciais. Enquanto sensores tradicionais têm alcance limitado, câmeras aliadas a métodos de visão computacional oferecem monitoramento em larga escala. Redes convolucionais superam técnicas baseadas em cor ou movimento, mas costumam ser pesadas; modelos mais leves, como o *YOLOv5*, equilibram eficiência e precisão.

O plano geral para abordar este projeto é aplicar o fine-tuning a um *YOLOv5* pré-treinado no *COCO* para especializá-lo nas classes “*Smoke*” e “*Fire*” usando o *D-Fire Dataset* (por volta de 21000 imagens). Posteriormente, será treinado um modelo *baseline* sem *augmentations* e, em seguida, técnicas *on-the-fly* (flips, rotações e ajustes fotométricos) serão introduzidas via *Albumentations/Ultralytics*.

Nas seções seguintes, será apresentado a definição do problema, os critérios de avaliação, a metodologia de *fine-tuning* e os resultados preliminares alcançados até este marco.

## 2. DEFINIÇÃO DO PROBLEMA

A tarefa é detectar e localizar regiões de “*Smoke*” e “*Fire*” em imagens estáticas, usando caixas delimitadoras para indicar cada ocorrência. Isso permite acionar alarmes ou notificações antes da propagação de chamas, reduzindo perdas e riscos. Fumaça e fogo exibem grande variabilidade de cor e textura (cinza-clara a quase transparente; amarelo a vermelho intenso), confundindo métodos de segmentação de cor em cenários com luzes artificiais, neblina ou baixa luminosidade. A falta de sequência temporal em imagens estáticas impede o uso de movimento para eliminar reflexos estáticos.

Para isso, utiliza-se o *D-Fire Dataset* (Venâncio et al., 2022) do *Kaggle*, com 21 527 imagens anotadas no formato *YOLO*: apenas fogo (1 164), apenas fumaça (5 867), ambas (4 658) e negativas (9 838), totalizando 26 557 caixas (11 865 fumaça; 14 692 fogo). As imagens vêm de *Internet*, simulações, câmeras de vigilância e composições sintéticas, garantindo variações de iluminação, escala e ambiente (industrial, residencial, natural). Mantém-se a divisão original (por volta de 80% treino, com validação interna, e 20% teste) para garantir avaliação justa.

Espera-se que o modelo *fine-tuned* atinja mais de 70 % de mAP@0.5 no teste, com Iou(Intersection over Union) médio >0,5 em ambas as classes. Qualitativamente, exibirá detecções corretas em cenários adversos e exemplos de falsos positivos (reflexos, objetos cor-de-fogo) e falsos negativos (chamas pequenas, fumaça translúcida).

## 3. ABORDAGEM TÉCNICA

A abordagem técnica fundamenta-se em *transfer learning*, empregando fine-tuning de uma variante *YOLOv5s* pré-treinada no *COCO Dataset*. Inicialmente, importa-se a implementação oficial em *PyTorch*, que fornece toda a arquitetura (*backbone*, *neck* e *head*) e

as rotinas de pré-processamento para carregamento de imagens e anotações no formato *YOLO* (coordenadas normalizadas de *bounding boxes*). A cabeça original do modelo—capaz de prever 80 categorias—é substituída por uma nova com apenas duas saídas, correspondentes às classes “*Smoke*” e “*Fire*”.

O *fine-tuning* utiliza o *D-Fire Dataset* em sua divisão oficial (treino, validação e teste). Em um primeiro experimento, não se aplica data augmentation. Com isso, as imagens são redimensionadas ao tamanho padrão do *YOLOv5*, normalizadas na faixa  $[0,1]$  e associadas às anotações de *bounding boxes* normalizadas. Os hiperparâmetros iniciais são: *learning rate* de 0,01 com *scheduler* baseado no desempenho de validação, *batch size* de 16 e 50 *epochs* de treinamento. Adota-se o otimizador SGD (*momentum* 0,937 e *weight decay* 0,0005), seguindo recomendações do *Ultralytics* para detecção em tempo real.

Após o *baseline*, introduzem-se transformações *on-the-fly* para aumentar a robustez: *flips* horizontais e rotações em  $90^\circ$ ,  $180^\circ$  e  $270^\circ$  via *Ultralytics*—que calculam automaticamente as coordenadas das caixas (por exemplo,  $(x, y) \rightarrow (y, W - x)$  no giro de  $90^\circ$ )—e ajustes fotométricos de brilho e contraste via *Albumentations*, sem alterar as posições das *bounding boxes*. Cada técnica é testada isoladamente para medir seu impacto nas curvas de perda e nas métricas de validação antes de compor uma única pipeline de *augmentations*.

Durante todo o processo, monitoram-se *classification loss*, *box loss* e métricas de validação (mAP@0.5, precisão e *recall*) a cada *epoch*, permitindo ajustes de *learning rate* ou número de *epochs* em caso de *overfitting* ou *underfitting*. Concluído o *fine-tuning*, realiza-se a inferência no conjunto de teste, armazenando as predições para análise quantitativa (mAP@0.5, IoU médio, precisão, *recall* e F1-score para cada classe) e qualitativa (figuras ilustrativas de detecções corretas, falsos positivos e falsos negativos, especialmente em condições de baixa luminosidade ou cenários visualmente complexos). Essa combinação de experimentos e métricas assegura que a abordagem seja reprodutível, mensurável e alinhada ao objetivo de produzir um detector confiável para aplicações de vigilância automatizada.

#### 4. RESULTADOS PRELIMINARES

Durante este marco, ajustou-se o *YOLOv5s* pré-treinado no *COCO* para as classes “*Smoke*” e “*Fire*” usando o *D-Fire Dataset*. Conforme demonstrado na Figura 1, as curvas de treinamento exibem uma convergência suave, em que o *box\_loss* declina de aproximadamente 0,07 para 0,04 e o *cls\_loss* de cerca de 0,014 para 0,004 ao longo de dez épocas, sem que haja espaçamento entre as curvas de treino e validação, o que evidencia boa generalização. Paralelamente, observa-se que *precision* e *recall* progridem de cerca de 0,50 para 0,70 e 0,65, respectivamente, enquanto o mAP@0.5 atinge o patamar de 0,70, e o mAP@0.5:0.95, métrica mais rigorosa, avança de 0,18 para 0,35. Esses resultados confirmam o sucesso do *fine-tuning* inicial e indicam que aprimoramentos adicionais, como a introdução de rotações e ajustes fotométricos, ou o refinamento de hiperparâmetros, poderão elevar ainda mais a performance do detector. A avaliação qualitativa também foi satisfatória, com o modelo sendo capaz de detectar a presença de fogo e fumaça em diferentes posições das

imagens de teste, uma amostra das imagens com as devidas *bounding boxes* pode ser encontrada no código referenciado no Anexo.

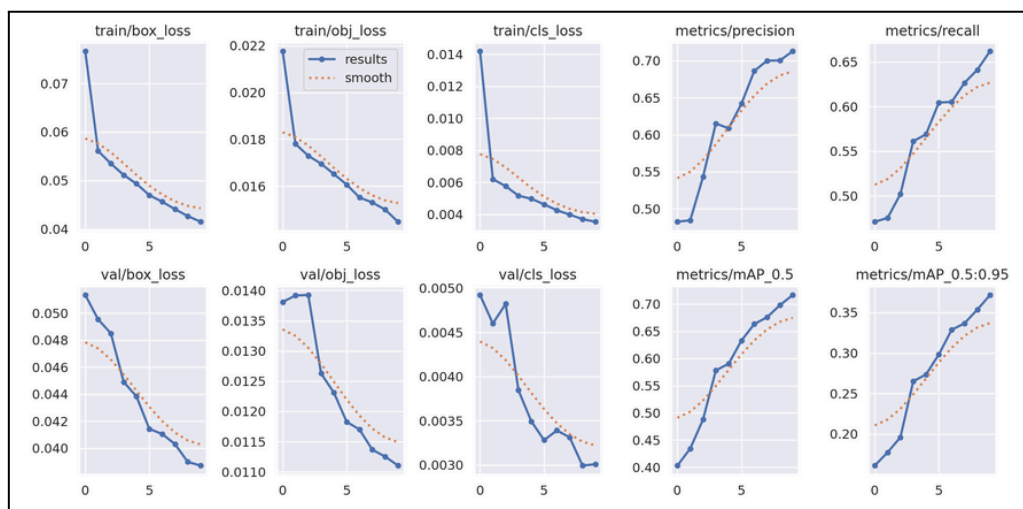


Figura 1: Métricas do modelo *YOLO* com *fine-tuning*



Figura 2: Reconhecimento de fogo e fumaça nas imagens do *dataset*.

Além disso, foi iniciada a implementação das operações de *data augmentation* nas imagens, começando pelo *flip* horizontal que foi executado nas imagens de treino do *Dataset*, e foram salvas em uma pasta. A implementação do *flip horizontal* está disponível nos Anexos, assim como, um exemplo dessa operação é demonstrado na Figura 3.



Figura 3: Aplicação de *flip* horizontal em imagem do *Dataset*

## 5. ANEXOS

A implementação inicial do modelo *YOLO* com *fine-tuning* está disponível em: <https://www.kaggle.com/code/joaomesa2/trabalhopi-fine-tuning>

Já a implementação dos *flips* horizontais no *dataset* está disponível em:

<https://www.kaggle.com/code/felipebonadia/trabalhopi-flip-horizontal>

## 6. REFERÊNCIAS

Pedro Vinícius Almeida Borges de Venâncio, Adriano Chaves Lisboa, Adriano Vilela Barbosa. "An automatic fire detection system based on deep convolutional neural networks for low-power, resource-constrained devices." *Neural Computing and Applications*, vol. 34, no. 18, 2022, pp. 15349–15368. DOI: 10.1007/s00521-022-07467-z.

J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.