



**IA901 - Processamento de Imagens e Reconhecimento de Padrões**

# **Reconhecimento de ação humana em imagens de veículos aéreos não tripulados (UAVs)**

**Discentes:** Débora Simões, Lucas Ueda e Wesna de Araujo

**Docentes:** Prof<sup>a</sup>. Letícia Rittner e a Prof<sup>a</sup>. Paula Dornhofer Paro Costa

# DESCRIÇÃO

- **OBJETIVO:** detectar pessoas e reconhecer o seu comportamento em imagens obtidas com drones.
- Exemplo de aplicação: planejamento de rotas *online* de UAV.
- Arquitetura adotada: YOLOv7.



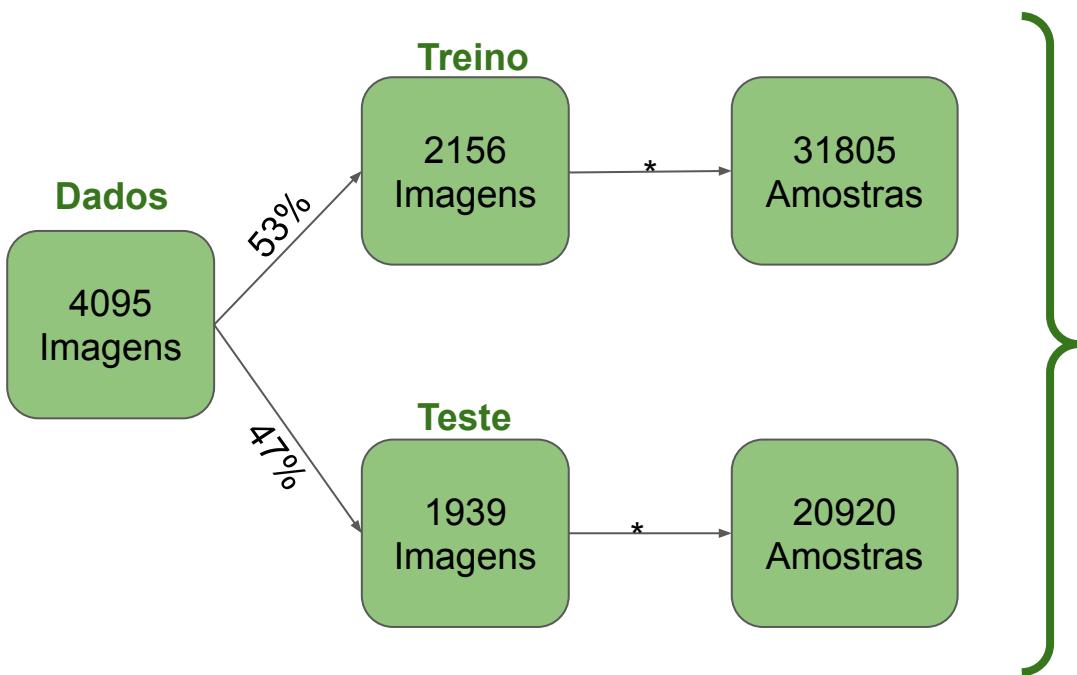
# BASE DE DADOS

- **Qual?** NTUT 4K Drone Photo Dataset for Human Detection
- **Descrição:** é formado por imagens de alta resolução extraídas de vídeos gravados por drones em Taiwan. Para cada imagem, estão disponíveis as coordenadas digitais que definem o bounding box em torno de uma pessoa detectada, bem como o rótulo que identifica a pose do indivíduo detectado.
- **Rótulos:** walk, stand, push, watch phone, baseball, sit e riding.
- **Dimensão:** 3840 x 2160 pixels
- **Quantidade:** 4095 imagens de drone



# ORGANIZAÇÃO DA BASE DE DADOS

- A seguir a organização do banco de dados bem como a quantidade de amostras por classe.

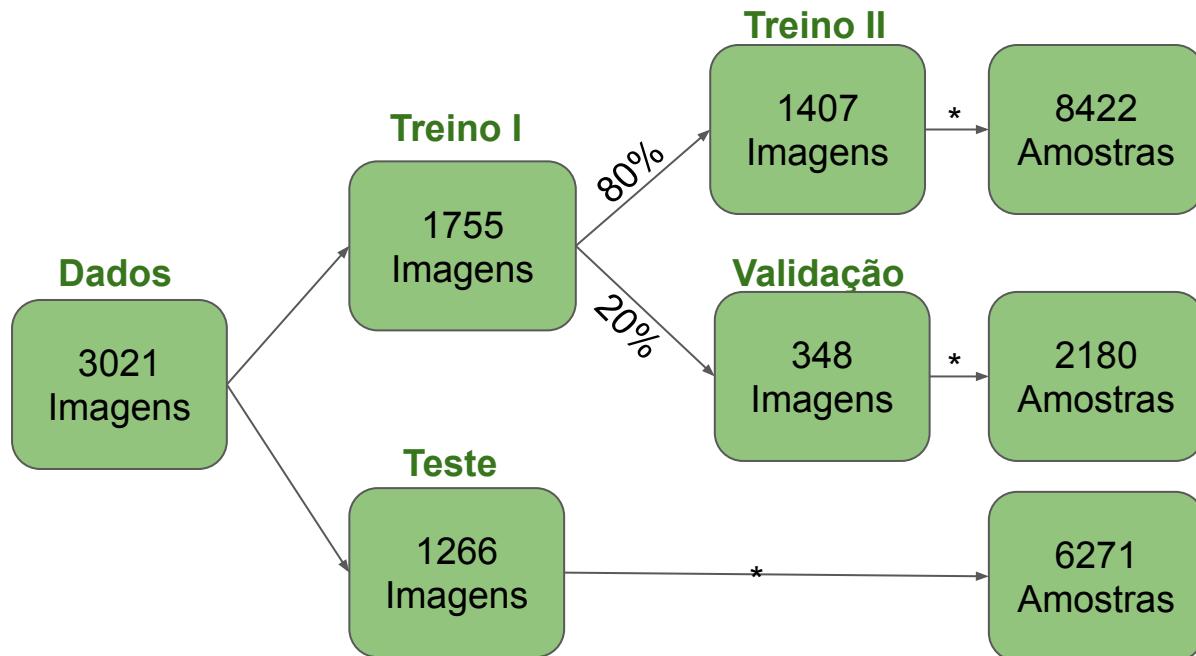


Classes	Treino	Teste
<b>walk (0)</b>	6155	4858
<b>stand (2)</b>	2551	1752
<b>push</b>	143	34
<b>watch phone</b>	167	0
<b>baseball</b>	118	0
<b>sit (3)</b>	424	481
<b>riding (1)</b>	2487	72
<b>outros</b>	19760	13723

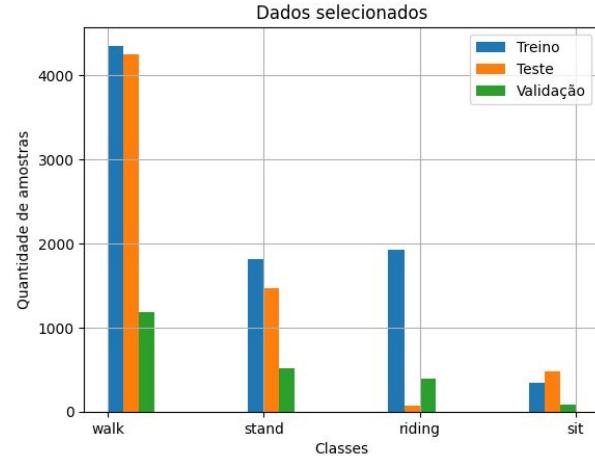
\*Mesma imagem pode ter mais de uma label.

# TREINO, VALIDAÇÃO E TESTE

- Considerando somente as classes walk, stand, riding e sit obteve-se a seguinte divisão de conjuntos



\*Mesma imagem pode ter mais de uma label.



Distribuição das amostras por classe.

- É importante mencionar, que como uma mesma imagem pode ser rotulada com mais de uma classe tomou-se o cuidado para que durante a divisão entre treino e validação (Treino I) imagens iguais pertencessem ao mesmo conjunto.

# METODOLOGIA

Organização dos dados



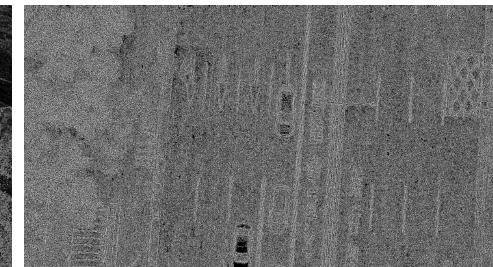
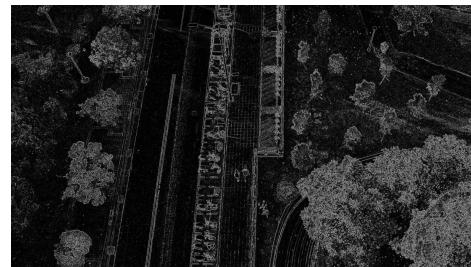
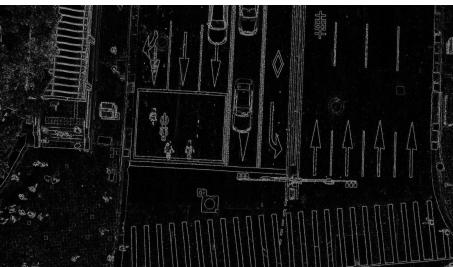
Pré-processamento dos dados

NC

Filtro de Sobel

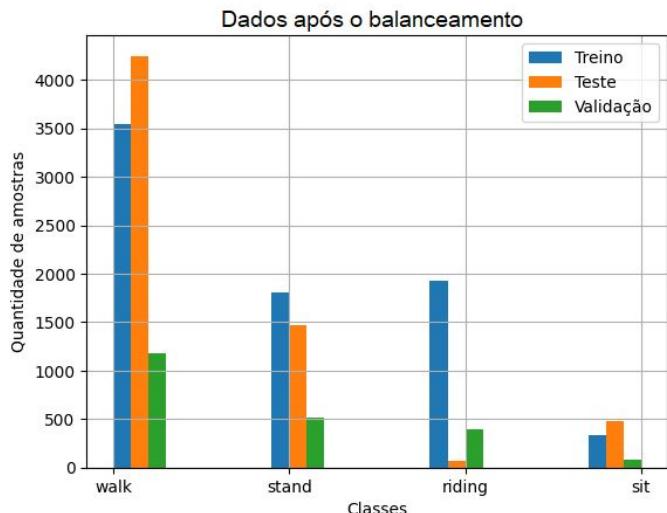
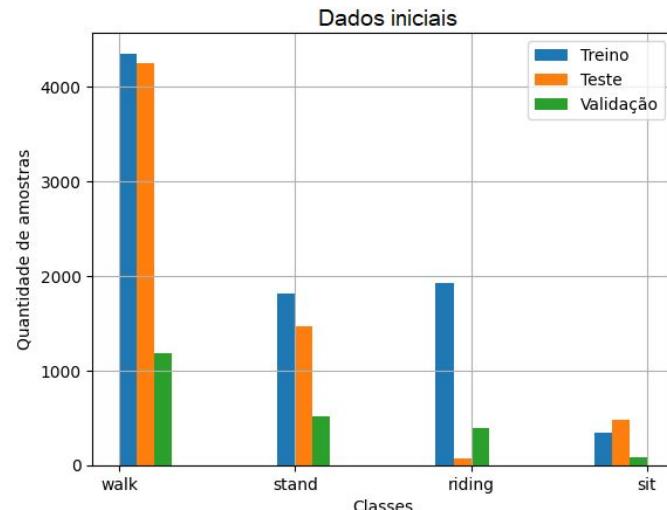
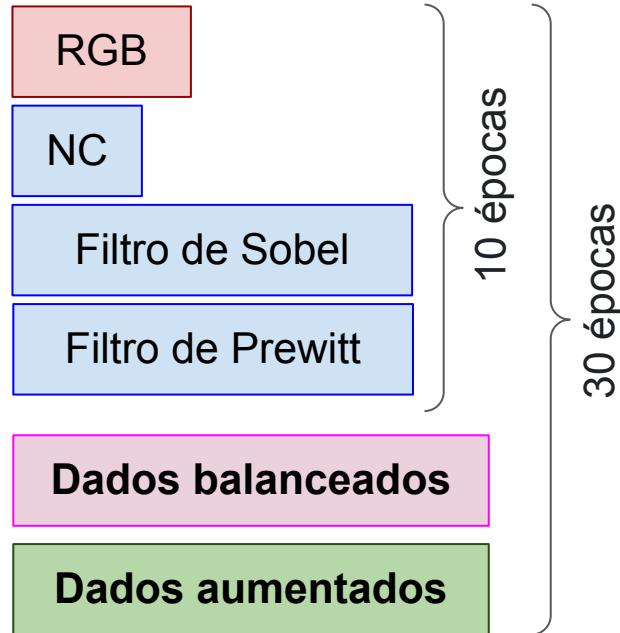
Filtro de Prewitt

Filtro de Laplace



# METODOLOGIA

## Treinamento do modelo



# METODOLOGIA

Dados aumentados

NC

Data augmentation

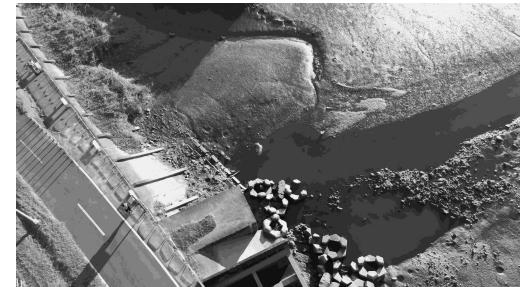
ColorJitter



RandomInvert



RandomPosterize



RandomSolarize



RandomAdjustSharpness



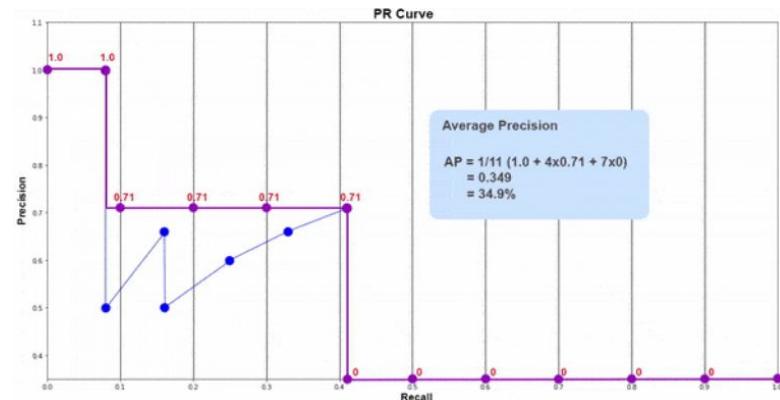
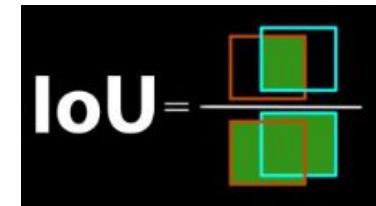
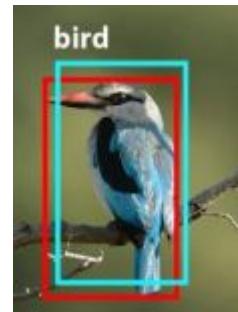
RandomEqualize



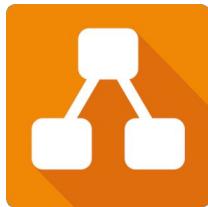
# METODOLOGIA

## Métricas avaliativas

- Precisão
- Recall
- mAP
  - mAP@.5: verdadeiros positivos  
→ limite IoU de 0.5;
  - mAP@.95: verdadeiros positivos  
→ limite IoU de 0.95.



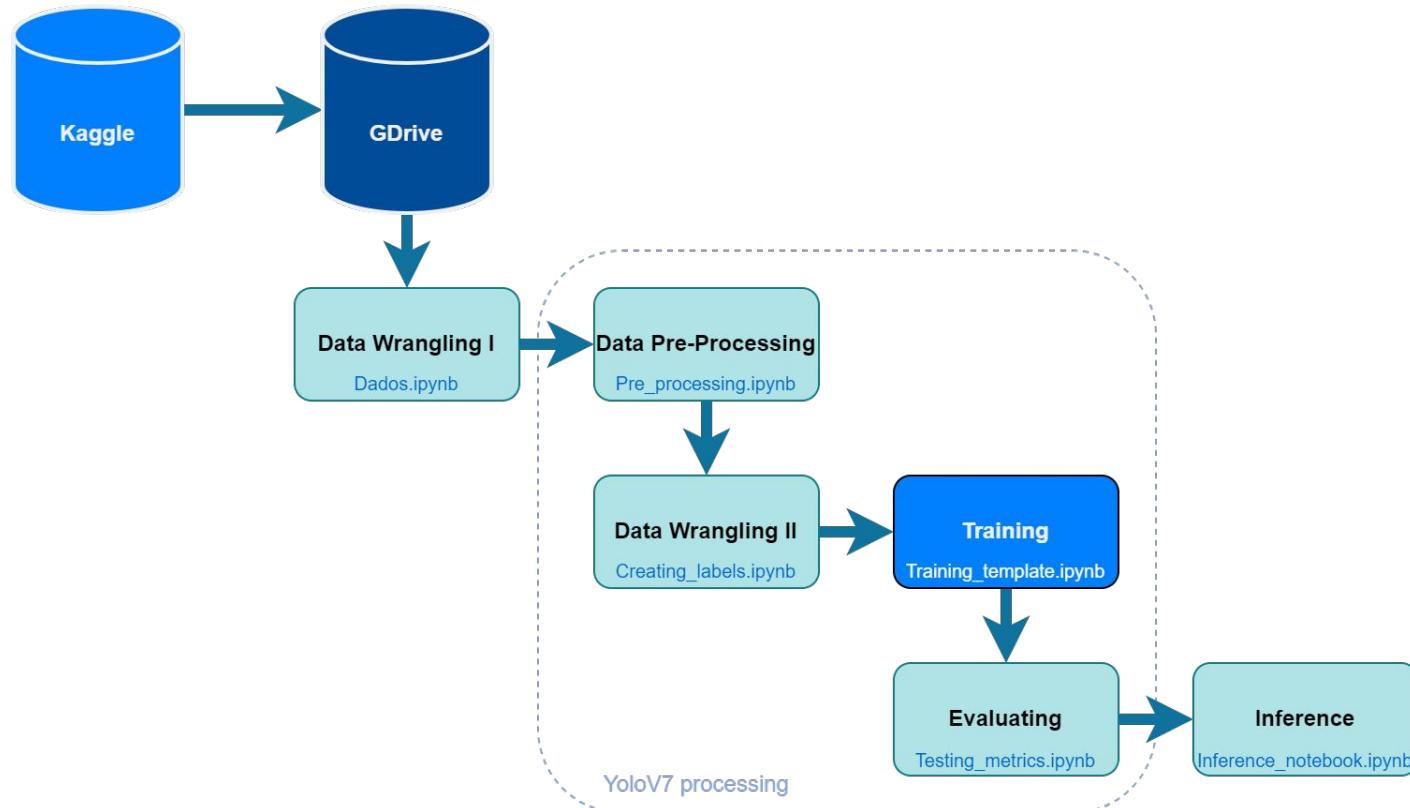
# PRINCIPAIS FERRAMENTAS



Algumas bibliotecas utilizadas:

- **OpenCV**: que auxiliou na visualização das imagens;
- **matplotlib.pyplot**: usada para plotar gráficos, com o intuito de verificar a distribuição de labels;
- **pandas**: utilizada para manipular dataframes;
- **numpy**: utilizada para operações matemáticas e cálculos envolvendo arrays provenientes dos dataframes;
- **torchvision**: usada para as transformações de imagens (data augmentation);
- **Scipy**: para a etapa de pré-processamento que necessitava do uso de convoluções;
- **os**: usada para operações do sistema, como acessar o conteúdo de pastas;
- **shutil.copy**: utilizada para copiar imagens de uma pasta para outra.

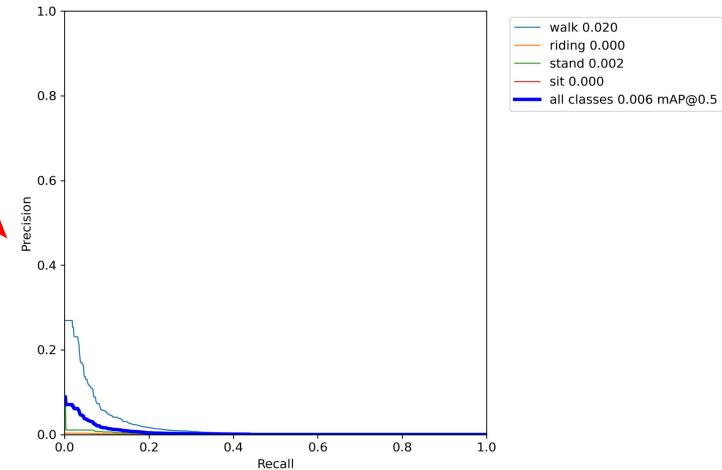
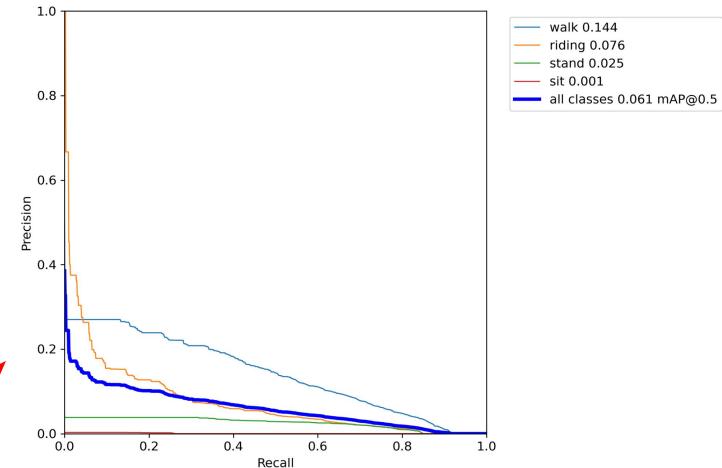
# WORKFLOW



# RESULTADOS E DISCUSSÃO

- Treinamento: 10 épocas

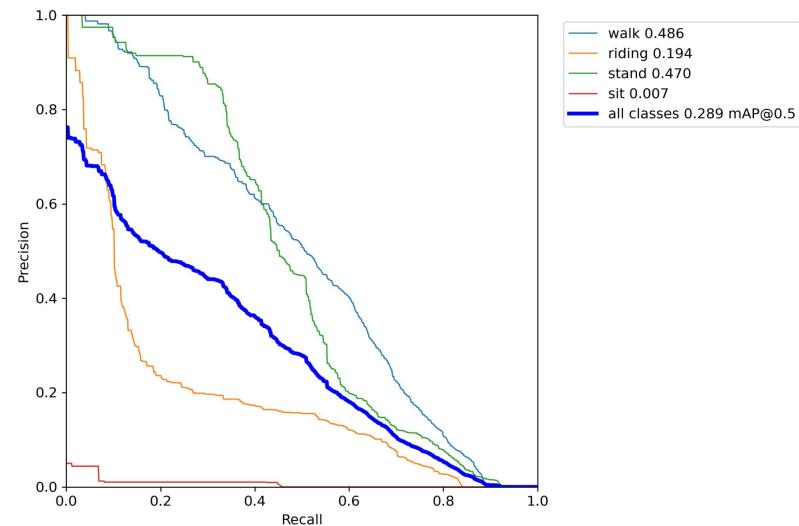
Processamento dos Dados	Precisão	Recall	mAP@0.5
Dados brutos filtrados	0.32	0,211	0.0625
Escala de Cinza	0.526	0.0168	0.00558
Filtro Sobel	0.555	0.131	0.0341
Filtro de Prewitt	0.277	0.073	0.0117



# RESULTADOS E DISCUSSÃO

- Treinamento: 30 épocas

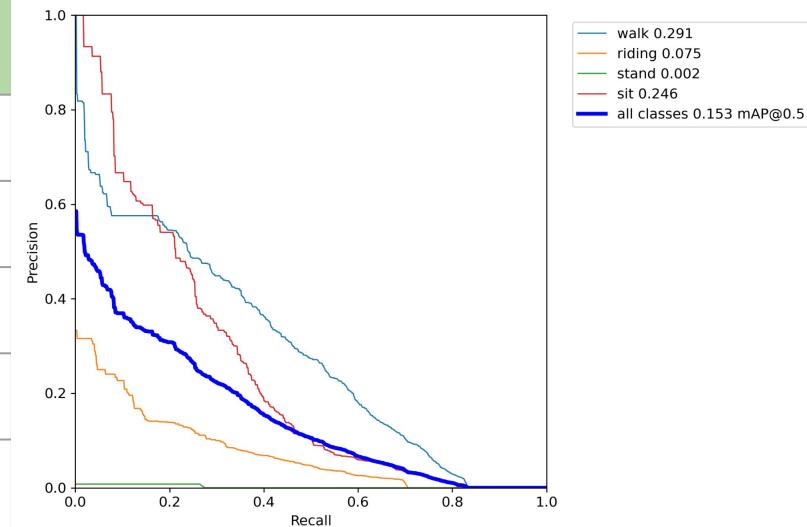
Processamento dos Dados	Precisão	Recall	mAP@0.5
Dados brutos filtrados	0.52	0.308	0.261
Escala de Cinza	0.564	0.307	0.289
Filtro Sobel	0.399	0.0824	0.0459
Filtro de Prewitt	0.391	0.172	0.0848



# RESULTADOS E DISCUSSÃO

- Treinamento: 30 épocas

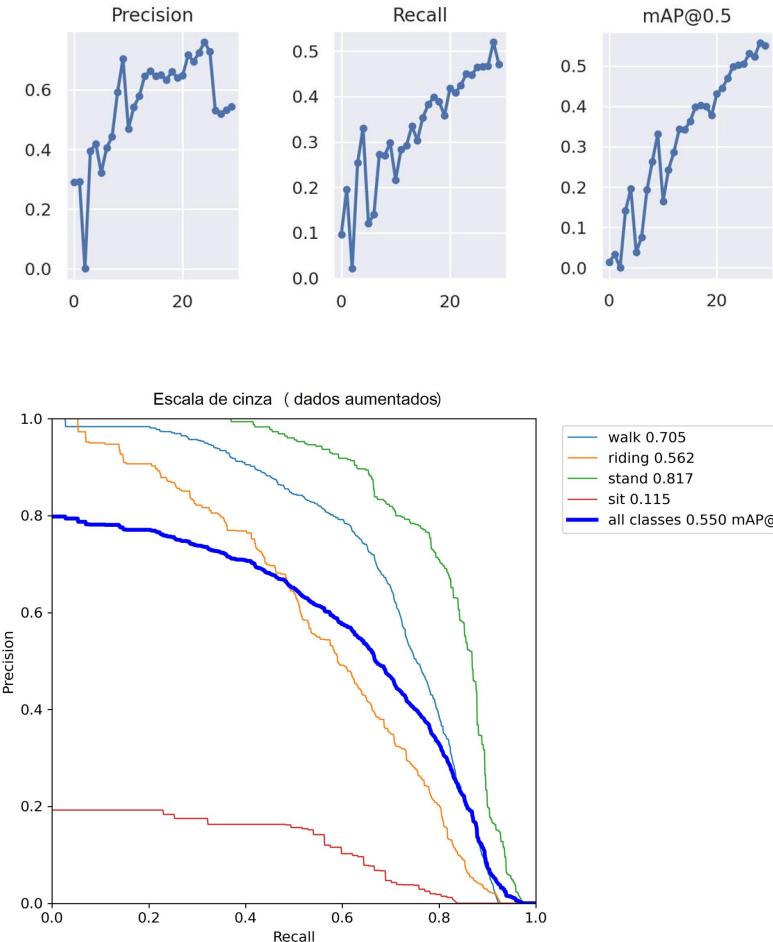
Processamento dos Dados	Precisão	Recall	mAP@0.5
Dados brutos filtrados	0.52	0.308	0.261
Escala de Cinza	0.564	0.307	0.289
Filtro Sobel	0.399	0.0824	0.0459
Filtro de Prewitt	0.391	0.172	0.0848
Escala de Cinza (dados平衡ados)	0.448	0.238	0.153



# RESULTADOS E DISCUSSÃO

- Treinamento: 30 épocas

Processamento dos Dados	Precisão	Recall	mAP@0.5
Dados brutos filtrados	0.52	0.308	0.261
Escala de Cinza	0.564	0.307	0.289
Filtro Sobel	0.399	0.0824	0.0459
Filtro de Prewitt	0.391	0.172	0.0848
Escala de Cinza (dados平衡ados)	0.448	0.238	0.153
Escala de Cinza (dados aumentados)	0.545	0.471	0.55



# RESULTADOS E DISCUSSÃO

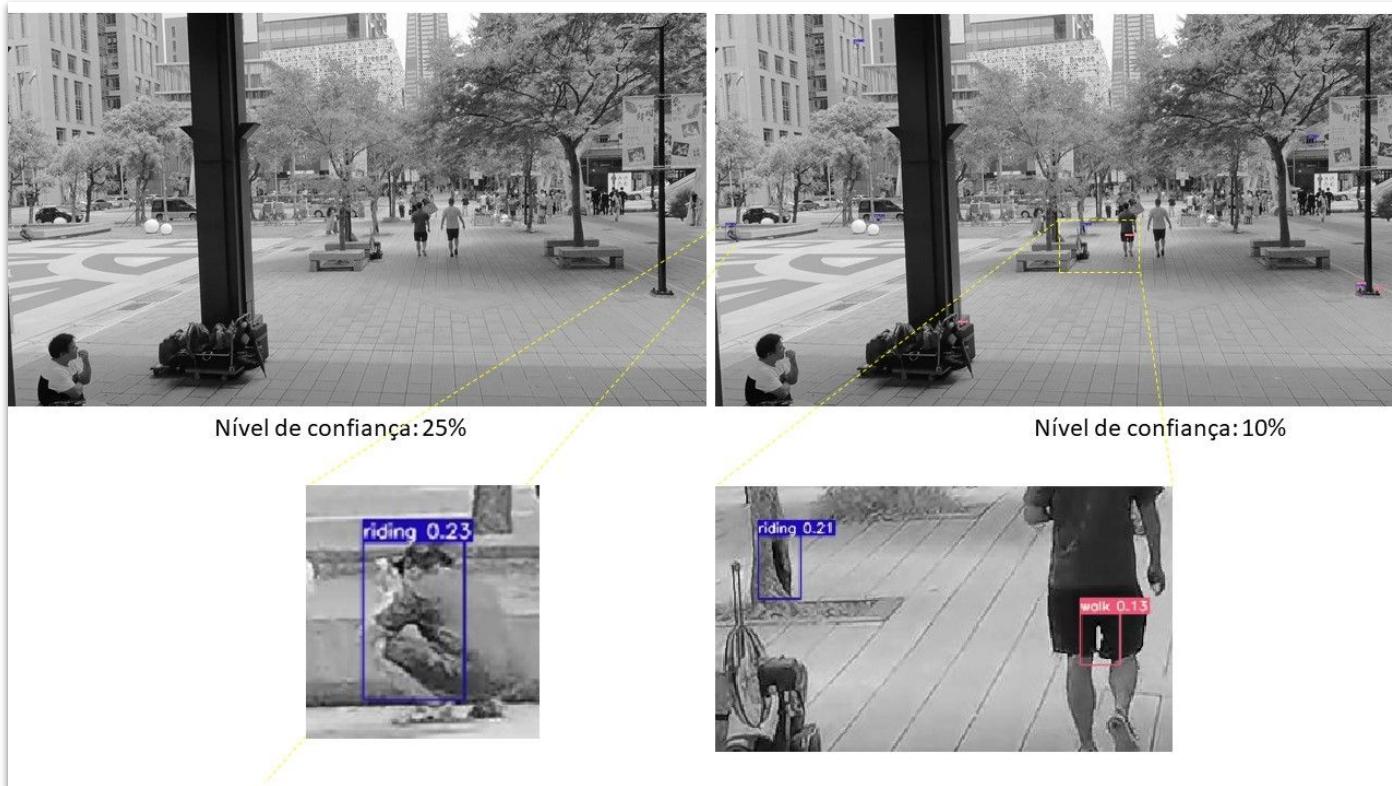
Ao fim de todos os experimentos, os modelos foram avaliados no conjunto de **teste** para 3 níveis de confiança de classificação: 25%, 10% e 0.1%.

Apesar do melhor modelo continuar sendo o NC+AUG, a performance caiu drasticamente em todos os experimentos.

Modelos	CONF	mAP .5 Test	mAP .95 Test
RGB	25,00%	3,37%	1,03%
	10,00%	4,83%	1,42%
	0,10%	6,03%	1,71%
NC	25,00%	4,52%	1,50%
	10,00%	5,48%	1,75%
	0,10%	6,94%	2,11%
NC+AUG	25,00%	<b>4,99%</b>	<b>1,80%</b>
	10,00%	<b>5,84%</b>	<b>2,04%</b>
	0,10%	<b>6,94%</b>	<b>2,31%</b>
NC+PREWITT	25,00%	0,22%	0,09%
	10,00%	0,46%	0,17%
	0,10%	5,58%	1,15%
NC+SOBEL	25,00%	NaN	NaN
	10,00%	NaN	NaN
	0,10%	0,74%	0,22%
NC+BALANCED	25,00%	2,07%	0,64%
	10,00%	3,41%	0,99%
	0,10%	4,36%	1,19%

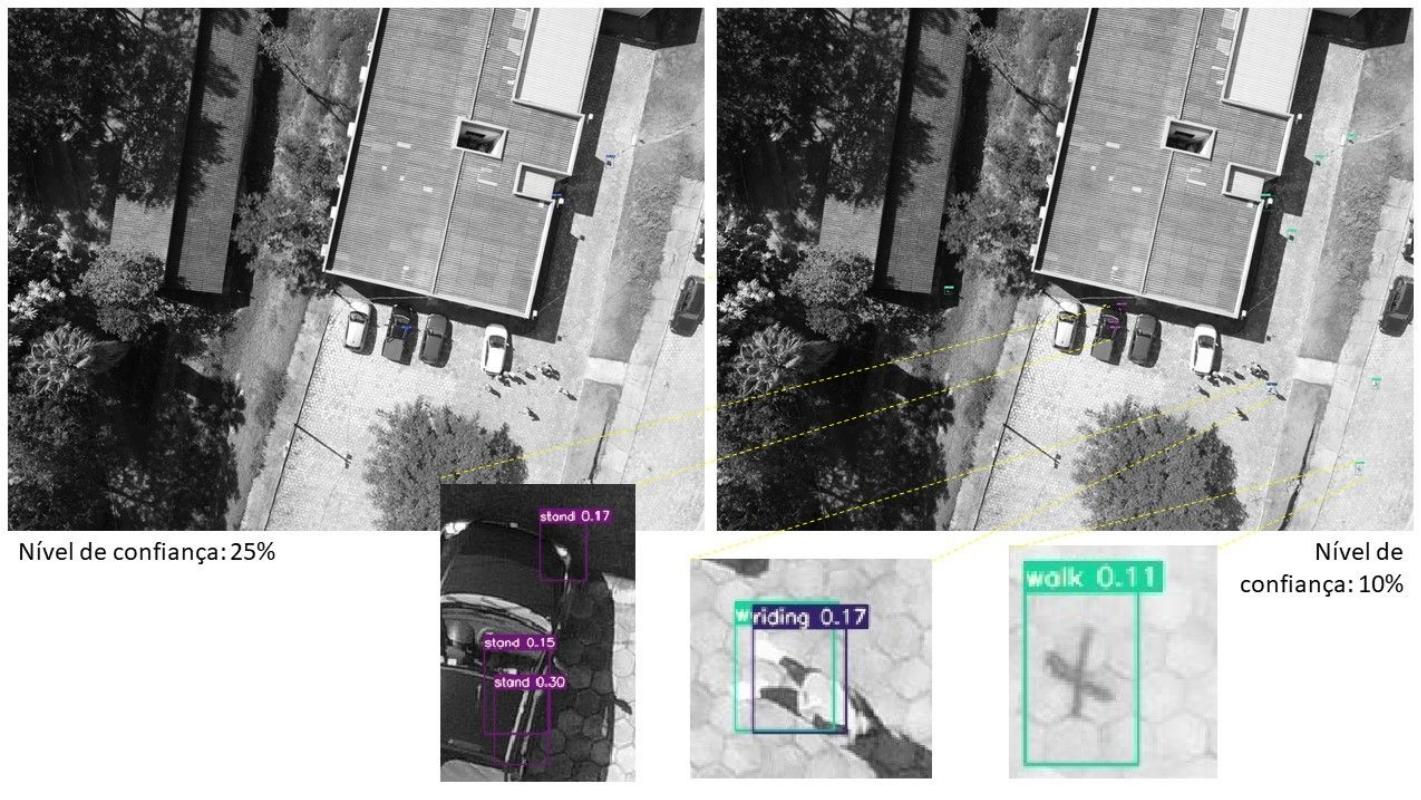
# RESULTADOS E DISCUSSÃO

- Inferência com imagens do dataset de teste (NTUT)



# RESULTADOS E DISCUSSÃO

- Inferência com imagens de Phantom 4 Pro



# CONCLUSÕES

- Utilizar as imagens em escala de cinza diminui o processamento ( $\frac{1}{3}$ ) além de manter a performance do modelo
- O filtro de Sobel e Prewitt não se mostraram bons pré-processamentos, bem como balancear as classes não trouxe resultados melhores
- Data augmentation foi a técnica mais efetiva para melhorar a performance do modelo

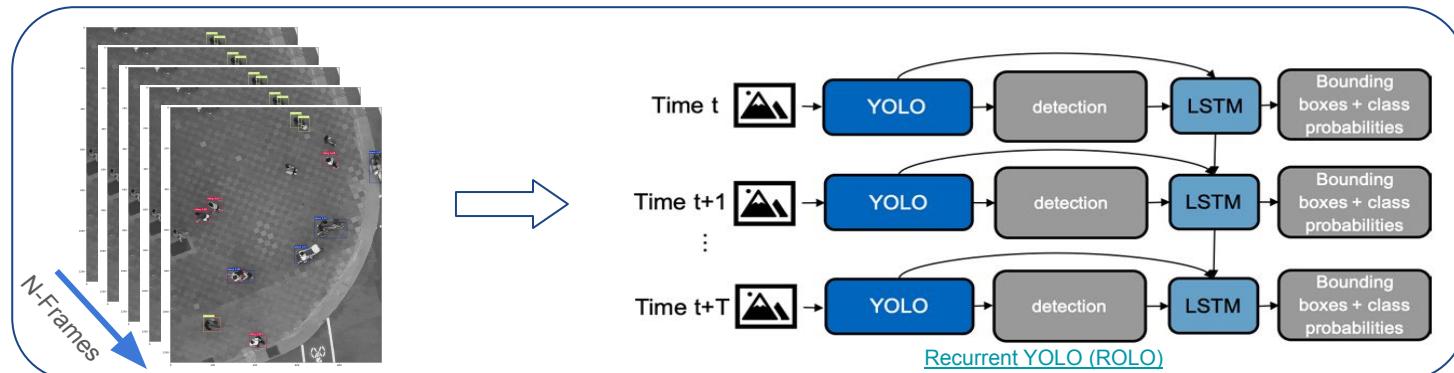
# TRABALHOS FUTUROS

- Analisar melhor o overfitting do modelo para possíveis melhorias marginais dos experimentos (como treinar por mais épocas e outras estratégias de data augmentation)
- Detectar pessoas e trackear sua movimentação



# TRABALHOS FUTUROS

- Analisar melhor o overfitting do modelo para possíveis melhorias marginais dos experimentos (como treinar por mais épocas e outras estratégias de data augmentation)
- Detectar pessoas e trackear sua movimentação
- Utilizar arquiteturas que levam em consideração informação temporal das imagens





# OBRIGADO!

Débora Simões, Lucas Ueda e Wesna de Araújo