

Presente! Sua sala de aula
inteligente.

Presente!

Sua sala de aula inteligente

Clayton Reis da Silva (clayton.reis@rnp.br)
Murilo Cruz Lopes (murilo.lopes@inatel.br)



Objetivo

- O objetivo do projeto é especificar e desenvolver um sistema de visão computacional que utiliza imagens de câmeras 5G, associado ao uso de dispositivos IoT, para uso em diversas aplicações no campus universitário, tais como detecção e reconhecimento de pessoas para controle de acesso, lista automática de presença em salas de aula, reconhecimento de comportamento, entre outras.

Executor: RNP

Co-Executores: Inatel

Prazo: 24 meses*

Previsão de término: 01/12/2026



Histórico

- Chamada Pública 2022 – MCTI/FINEP/FNDCT: Desenvolvimento de Aplicações em Redes 5G Privadas
- Linha Temática: Cidades Inteligentes
- Instituições Participantes:

“Nos últimos anos, vem-se observando um crescimento substancial do uso de inteligência artificial para detecção e identificação de pessoas em tempo real para monitoramento e segurança.”

(José Ferreira Rezende - Coordenador do Projeto)



Inatel



Equipe

Contamos com uma equipe multidisciplinar formada por mestres e doutores, que atuam como Coordenadores de P&D, Coordenadores Projetos, Analistas Administrativo e de Negócios e Pesquisadores, nas três instituições participantes: RNP, Inatel e UFRJ. Essa colaboração fortalece nosso trabalho e garante uma abordagem abrangente e especializada em todas as etapas do projeto.



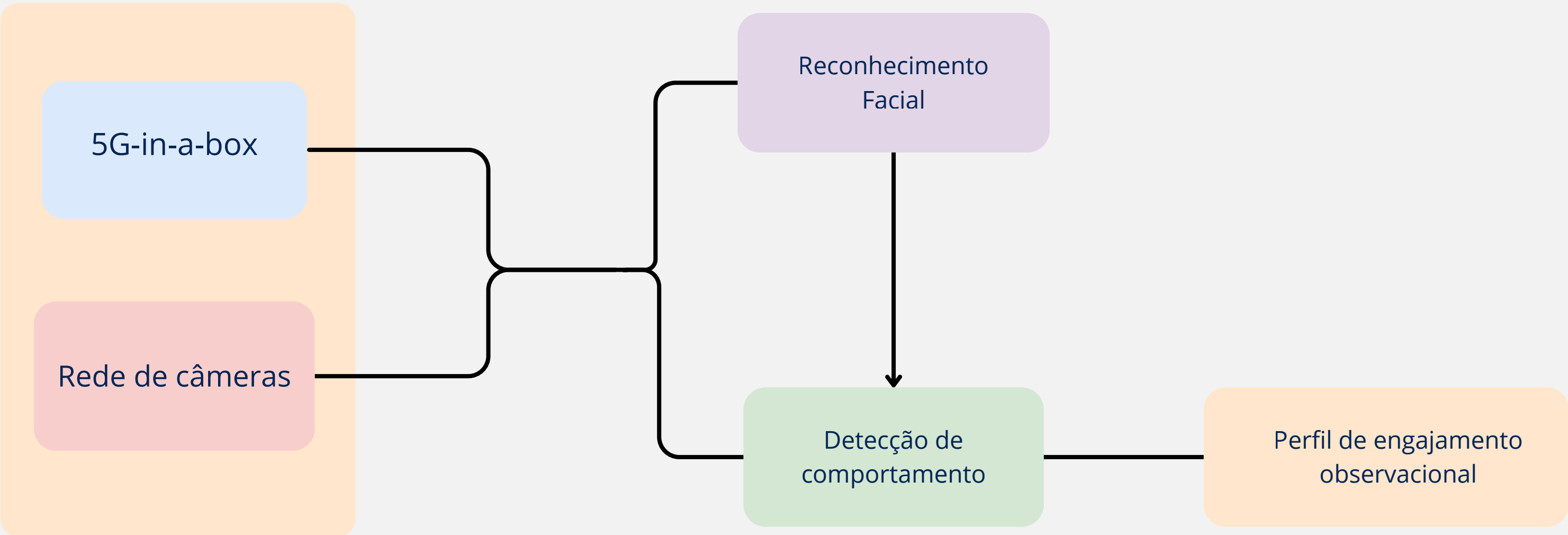
Desafio

O grande desafio do projeto é medir o engajamento dos alunos em tempo real, identificando os comportamentos observados durante as aulas de forma precisa e eficiente.



Proposta de solução

Criar um sistema de visão computacional conectado a uma rede privada 5G para estudar o engajamento observacional de alunos em sala de aula no Inatel

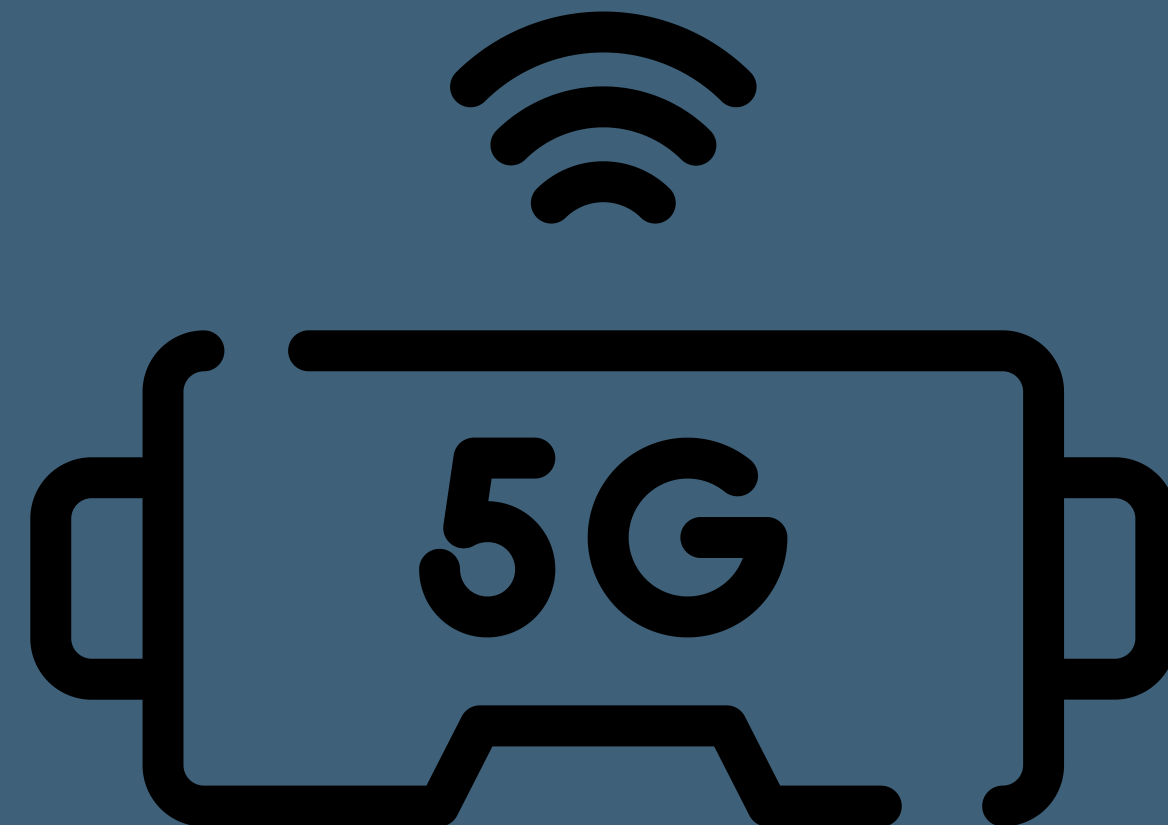


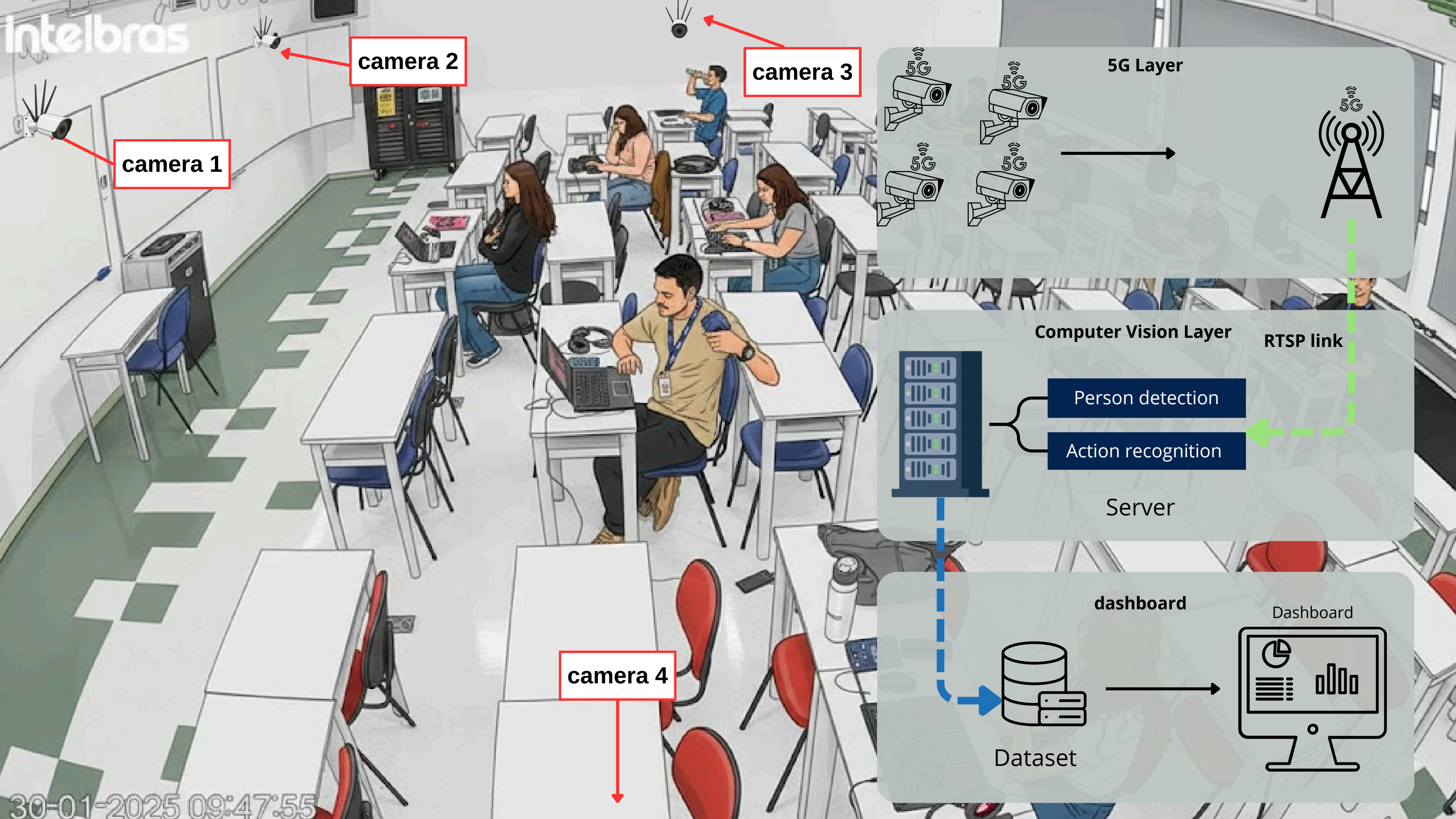
Visão 5G



Características de um sistema em Visão 5G

- Modular e desagregável;
- Uso de protocolos avançados de transmissão de vídeo;
- Sistema de comunicação entre câmeras.
- Infraestrutura computacional baseada em microsserviços;
- Análises em tempo real;



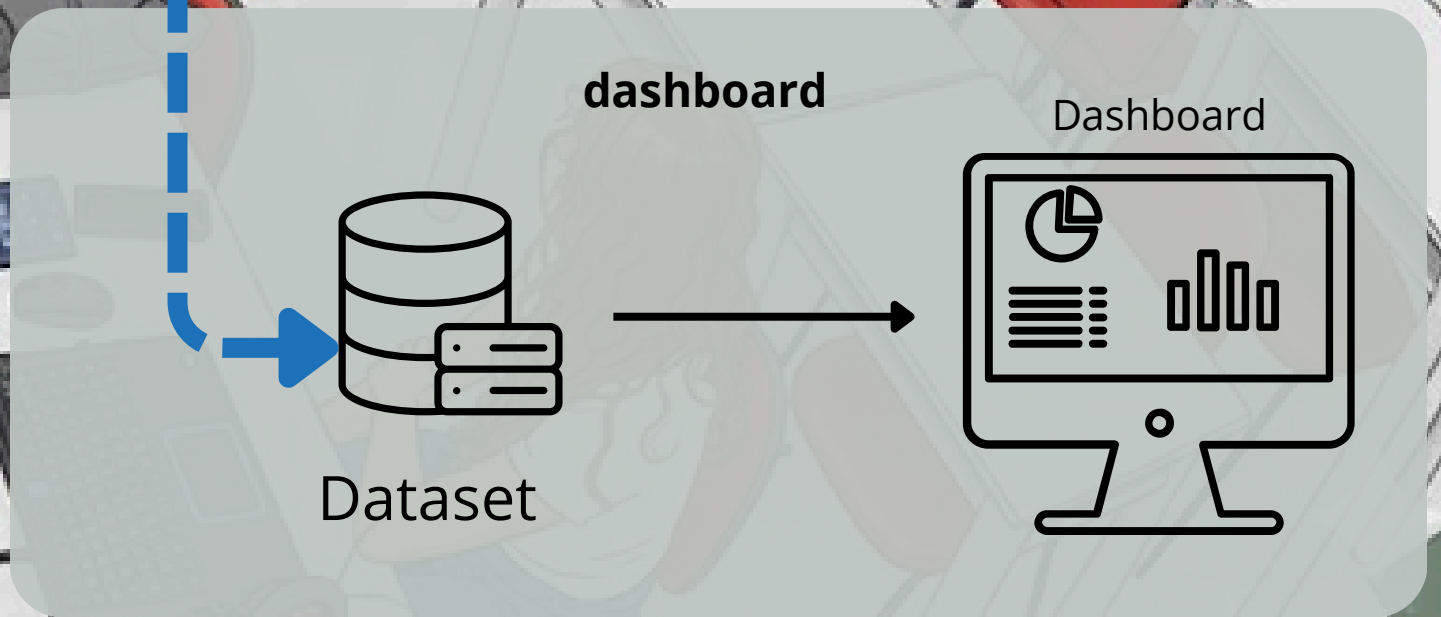
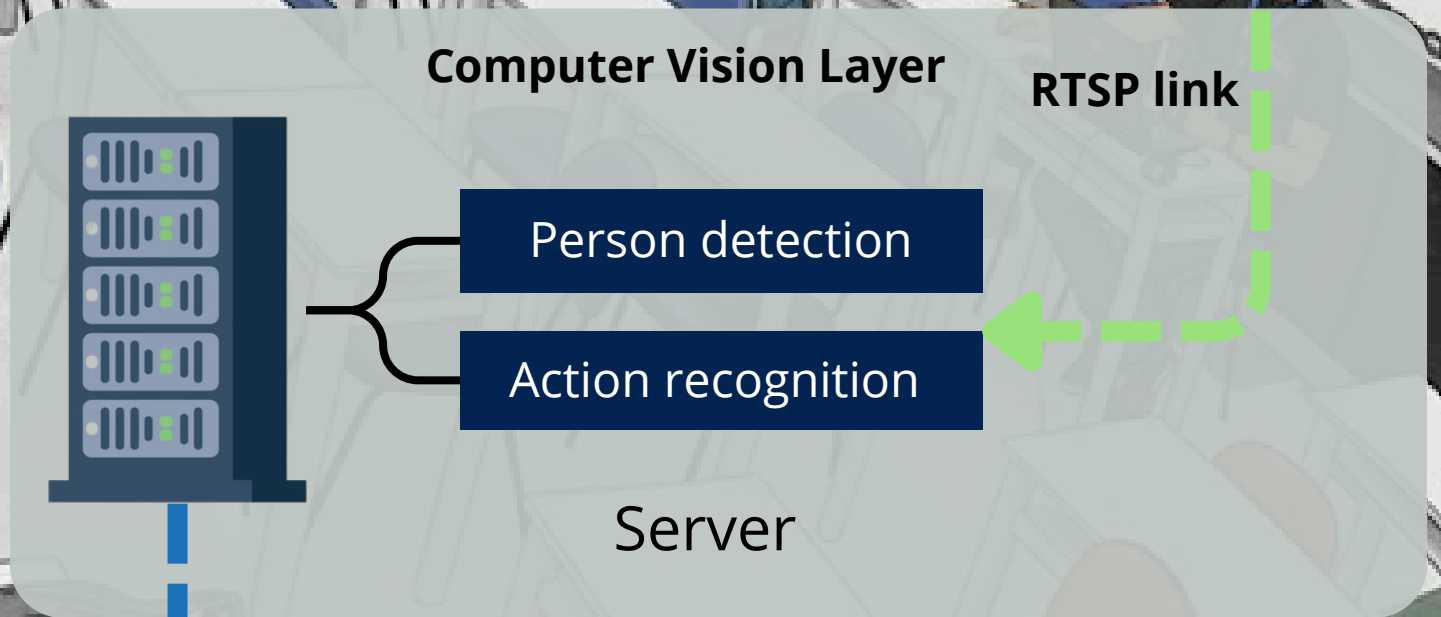
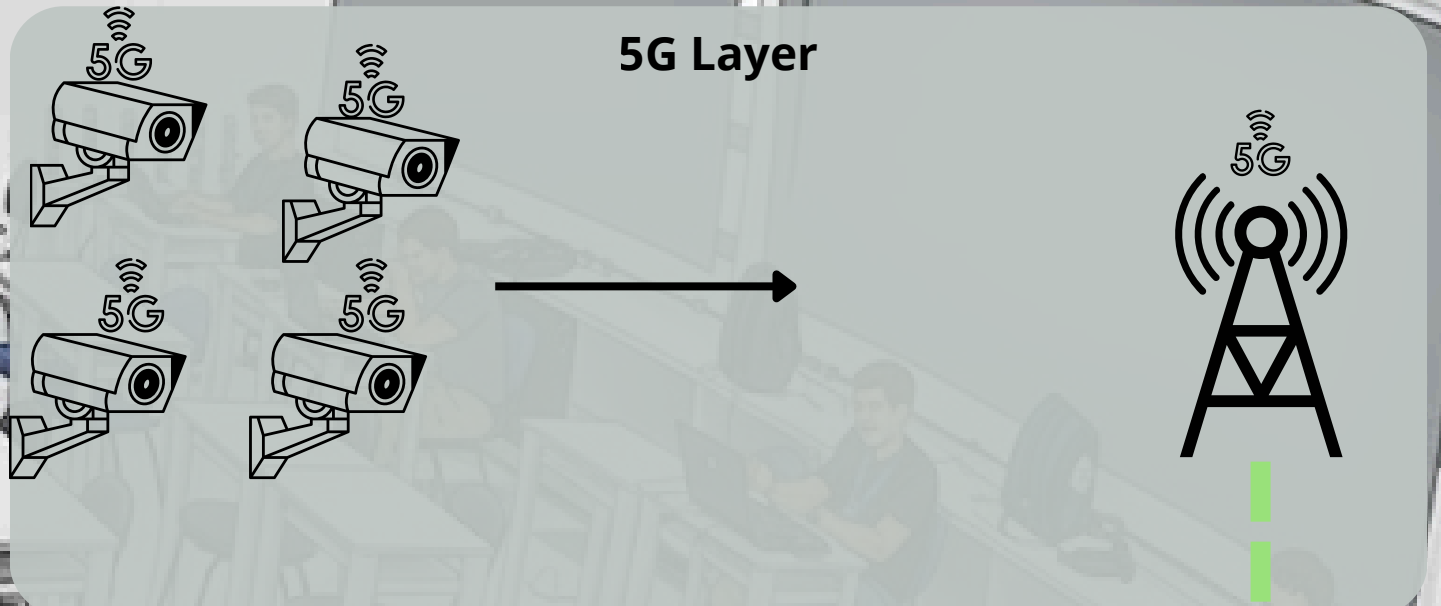


camera 1

camera 2

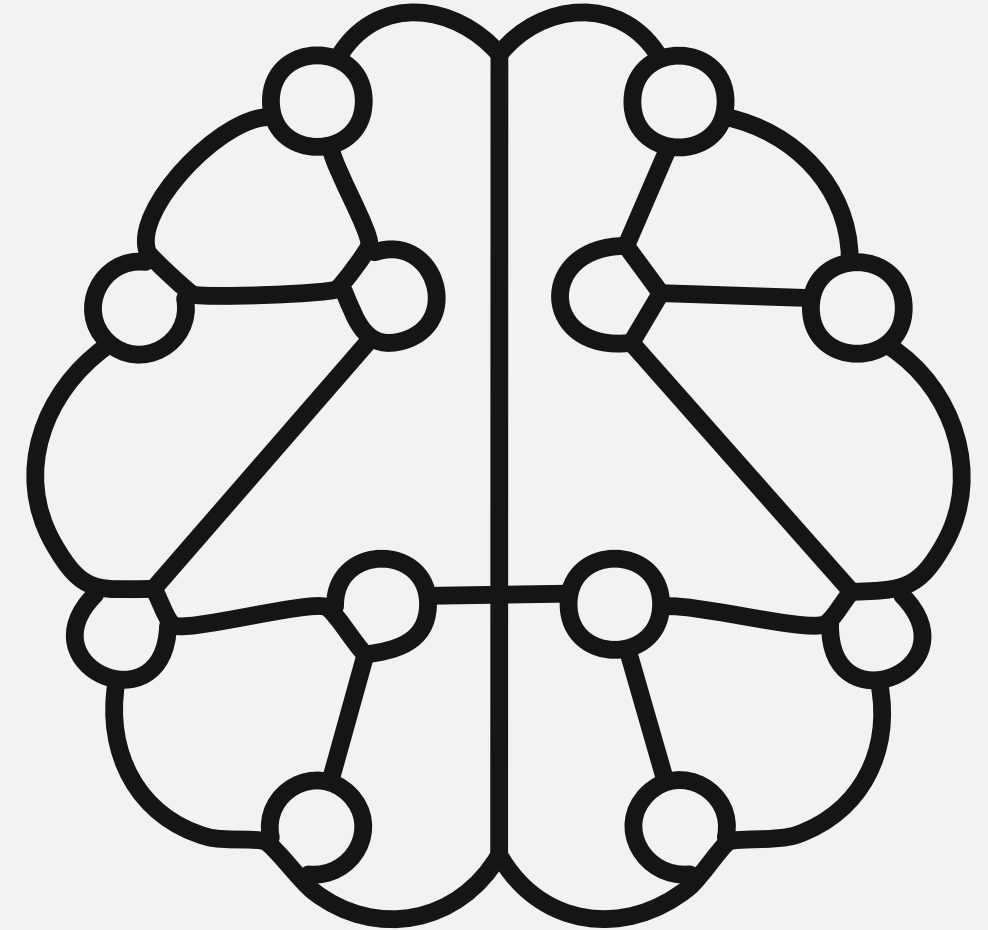
camera 3

camera 4



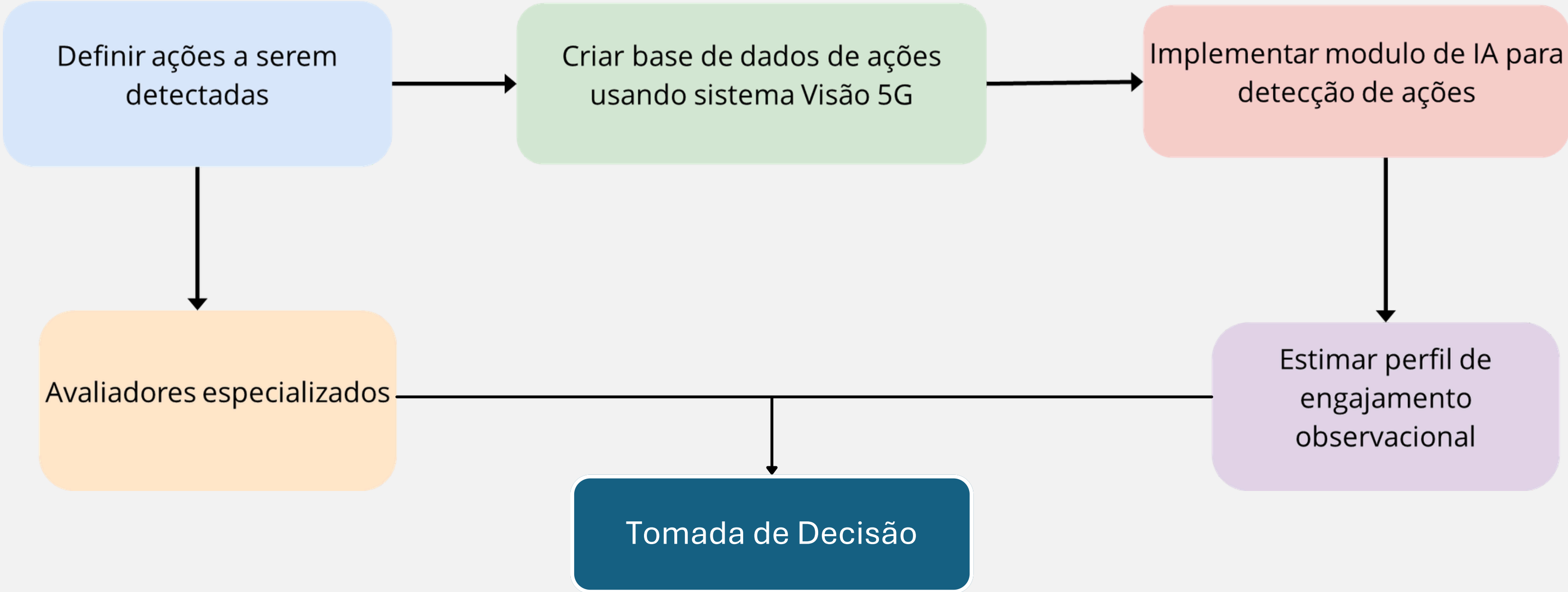
Etapas do Treinamento

- Definição das ações
- Anotação dos clips
- Divisão da Base de dados
- Definição dos modelos e parâmetros
- Análise dos Resultados



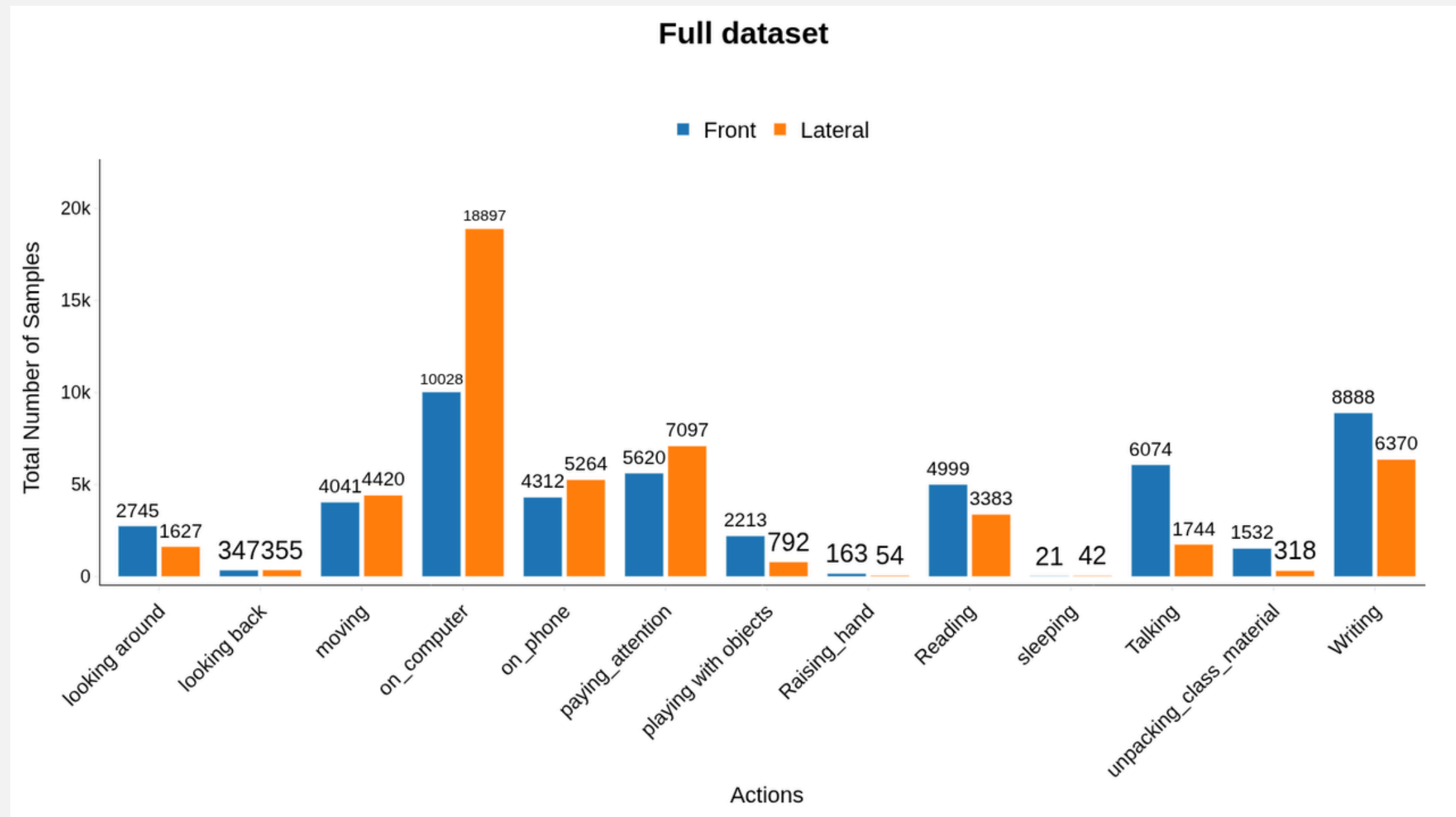
Detecção de comportamento e engajamento observacional

Ações associadas com o envolvimento em atividades escolares



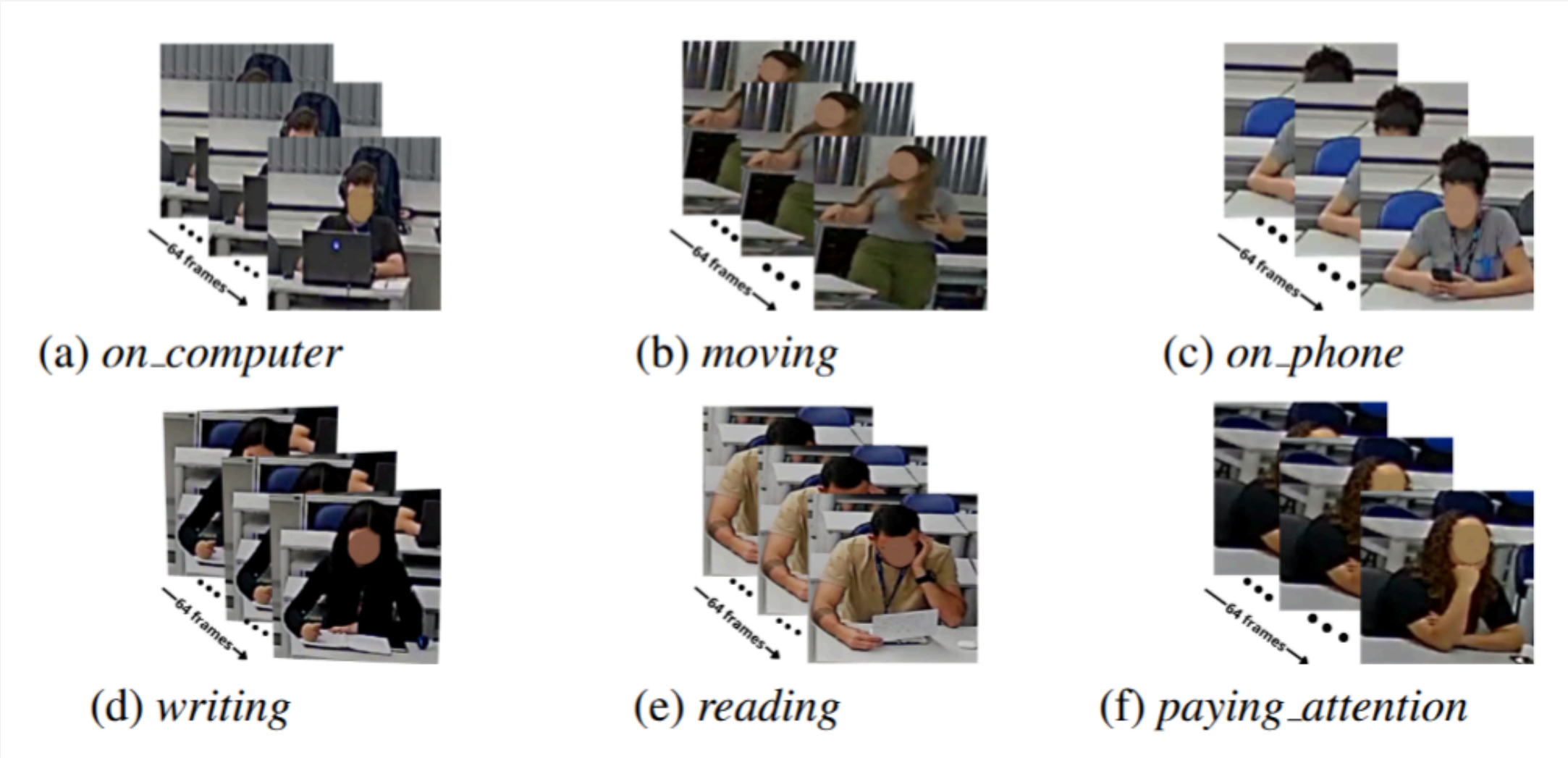
Estatísticas do Dataset

A anotação sistemática de 468 horas de dados longitudinais de sala de aula resultou em um conjunto de dados final com mais de 96.000 vídeos padronizados de 2,13 segundos. Este total inclui 45.663 perspectivas laterais e 50.983 perspectivas frontais.



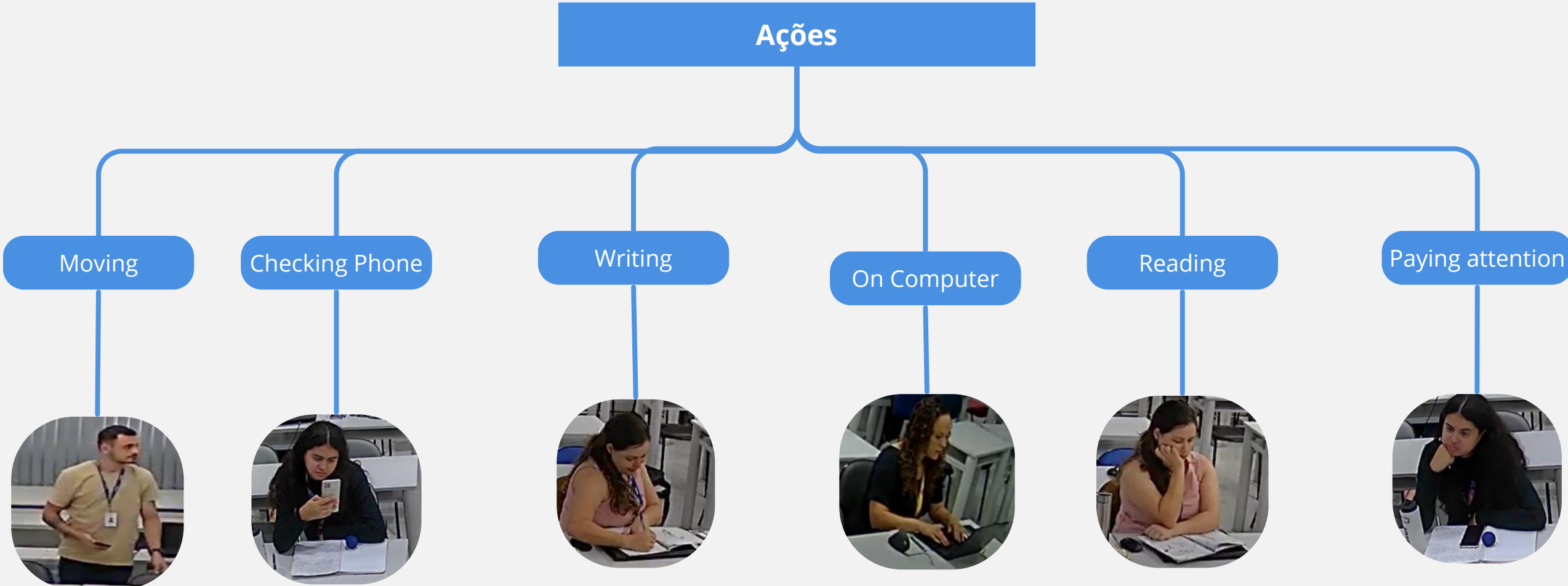
Detecção de comportamento e engajamento observacional

Conjunto de ações extraídas a partir de protocolos de engajamento observacional



Detecção de comportamento e engajamento observacional

Conjunto de ações extraídas a partir de protocolos de engajamento observacional



Divisão do Dataset

- Divisão da base de dados para as ações selecionadas
- Divisão da base de dados por aula para evitar vazamento de dados

Classe	Frontal Cameras			Lateral Cameras		
	Train	Val.	Test	Train	Val.	Test
on_phone	2779	500	500	2802	508	500
paying_attention	2637	527	514	2995	557	579
Reading	2223	524	508	1578	500	530
Writing	2211	506	527	2692	500	500
moving	2187	517	514	2597	500	511
on_computer	1920	504	568	2988	533	500
Total	13957	3078	3131	15652	3098	3120

Redes e Parâmetros

Conjunto de redes testadas para classificação de video clip
Conjunto de parâmetros para aplicação de GridSearch

Parameter	Value
Model	mvit_v2, mvit_v1, slowFast_r101, X3D_m, swin3D_s
Number of epochs	20
Batch size	32, 64
Learning rate	2×10^{-5} , 5×10^{-5}
Dropout	0.2, 0.4
Optimizer	AdamW
Weight decay	0.05
Scheduler	StepLR / Cosine Annealing
Warmup	3, 5 epochs
Loss function	Cross-Entropy
Number of classes	6
Data augmentation	RandomCrop, Horizontal Flip
Mixup (α)	0.4, 0.8
Seed	42

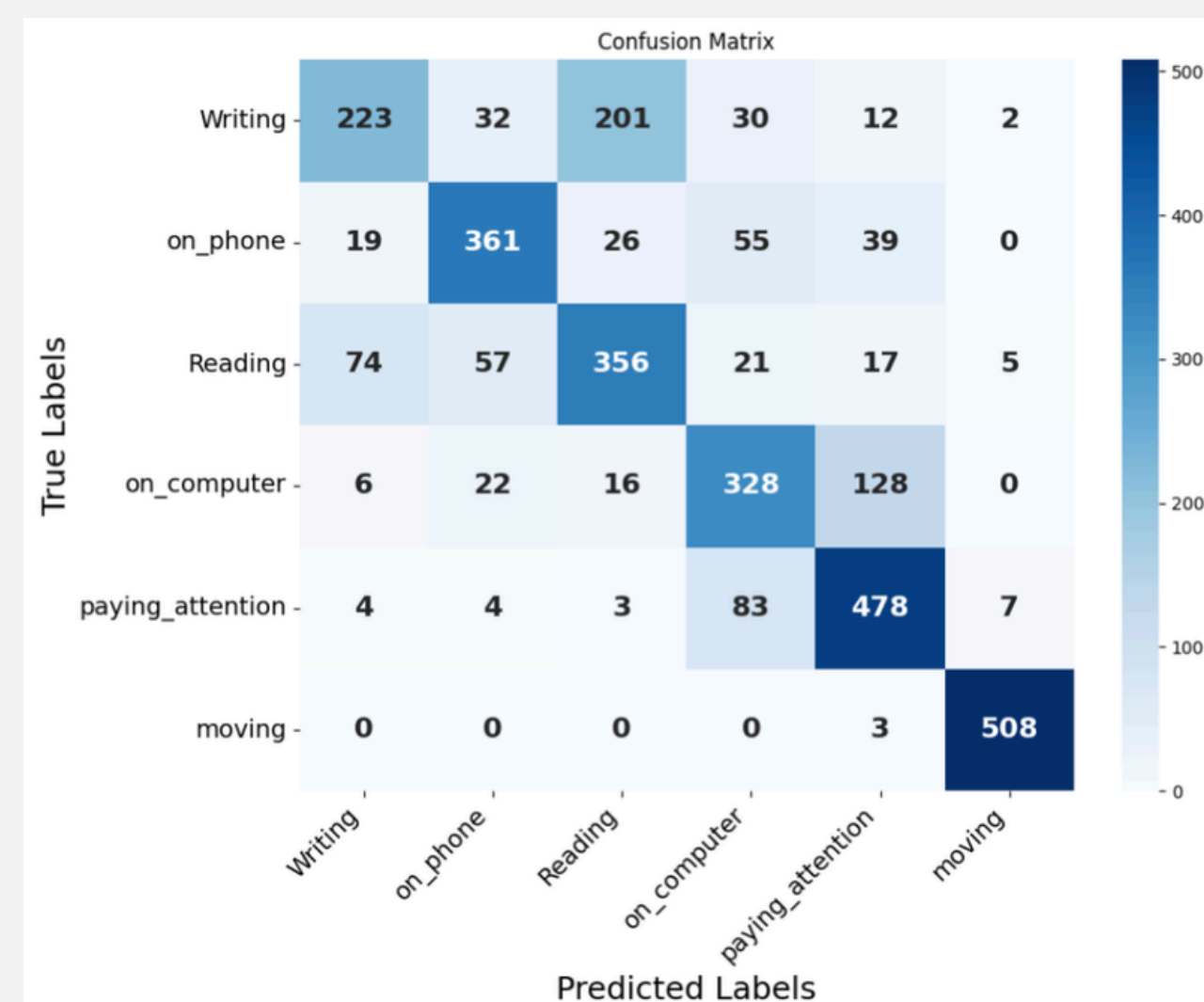
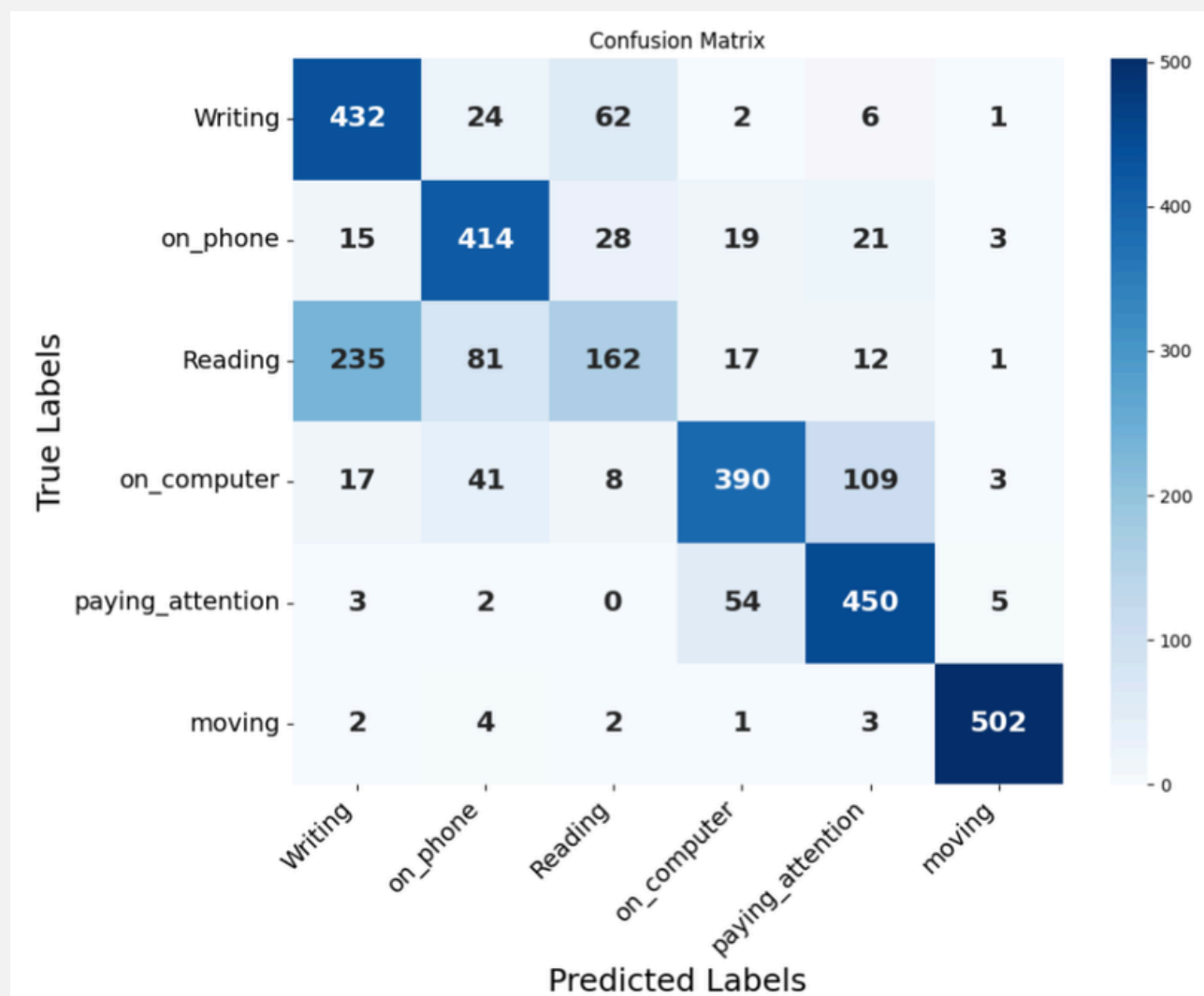
Resultados

Câmera Frontal

Algorithm	Accuracy (%)	F1-score (%)	AUC ROC (%)
MViT_v1_b	72.78	71.28	83.77
<i>MViT_v2_s</i>	75.04	73.74	85.05
<i>Swin3d_s</i>	73.58	71.92	84.18
<i>Slowfast_r101</i>	70.74	68.74	82.56
<i>X3D_M</i>	70.90	69.97	82.52

Câmera Lateral

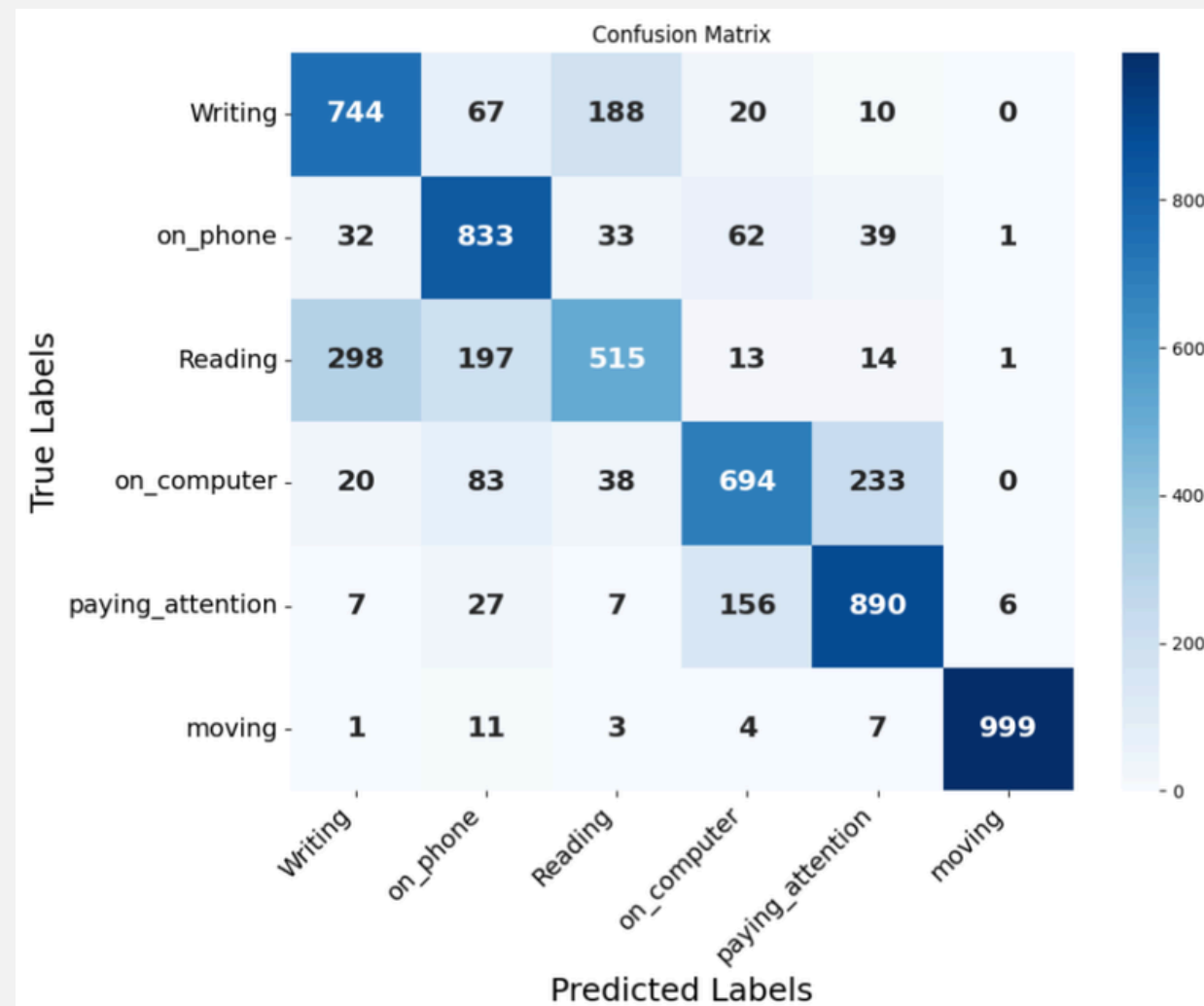
Algorithm	Acc. (%)	F1 (%)	AUC ROC (%)
MViT_v1_b	70.60	70.08	82.13
<i>MViT_v2_s</i>	72.24	71.64	83.17
<i>Swin3d_s</i>	70.57	69.87	82.16
<i>Slowfast_r101</i>	66.69	65.25	79.78
<i>X3D_M</i>	63.65	61.18	77.85



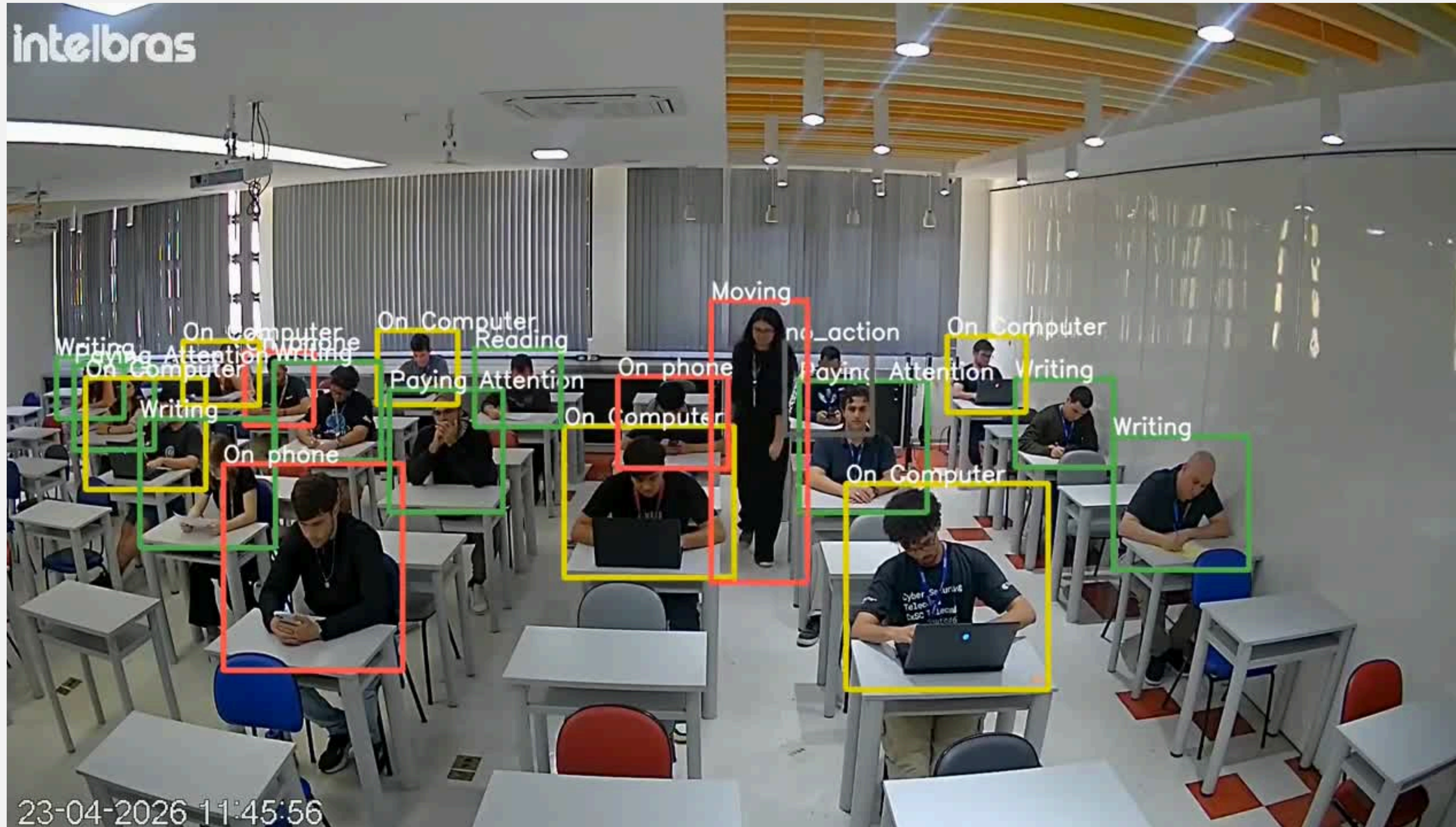
Resultados

Todas as câmeras

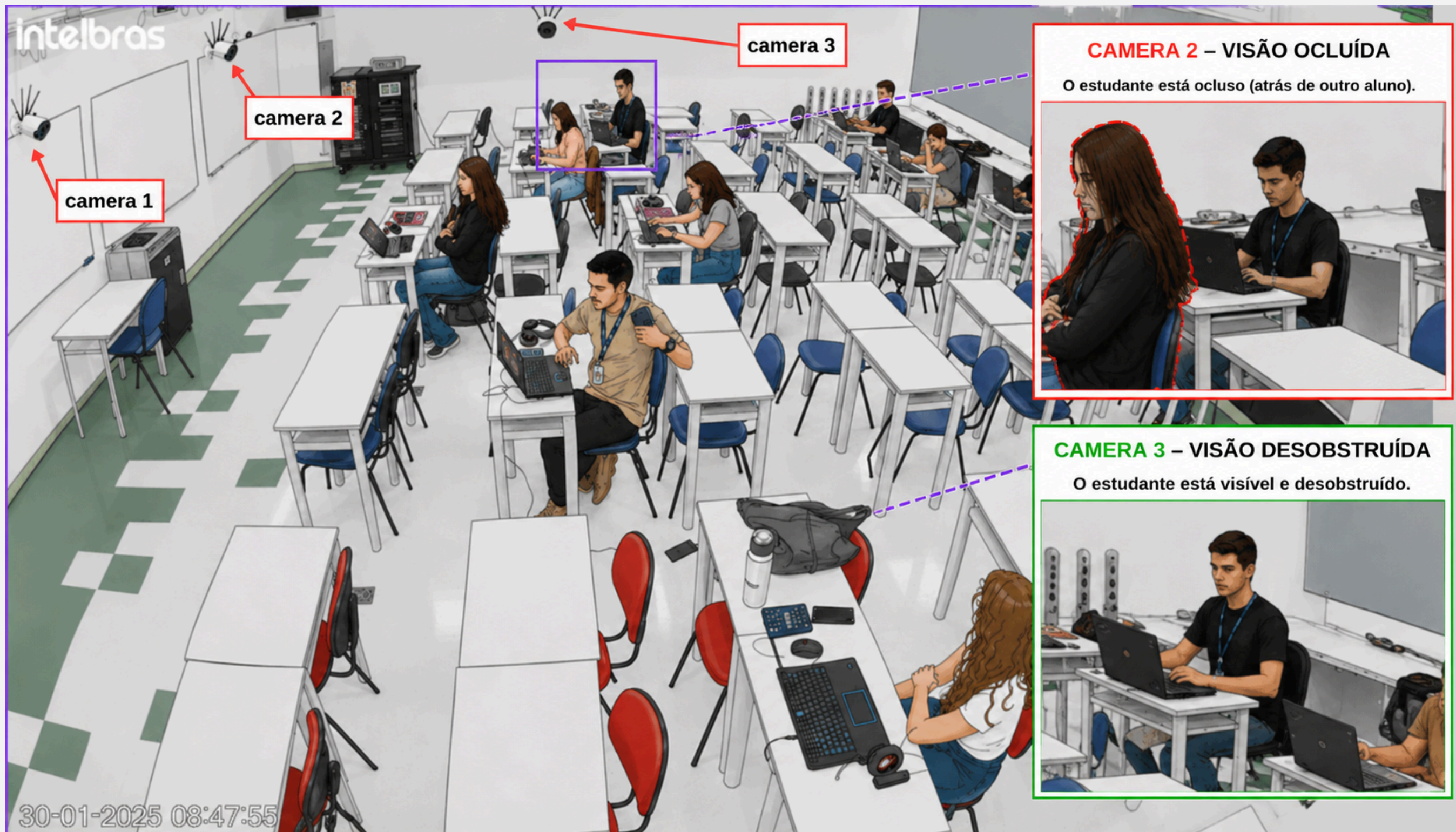
Algorithm	Acc. (%)	F1 (%)	AUC ROC(%)
MViT_v1_b	72.42	72.13	83.48
<i>MViT_v2_s</i>	74.78	74.43	84.91
<i>Swin3d_s</i>	71.73	71.26	83.08
<i>Slowfast_r101</i>	69.30	68.46	81.63
<i>X3D_M</i>	67.38	67.46	80.48



Primeiros resultados do modelo de reconhecimento de Ações

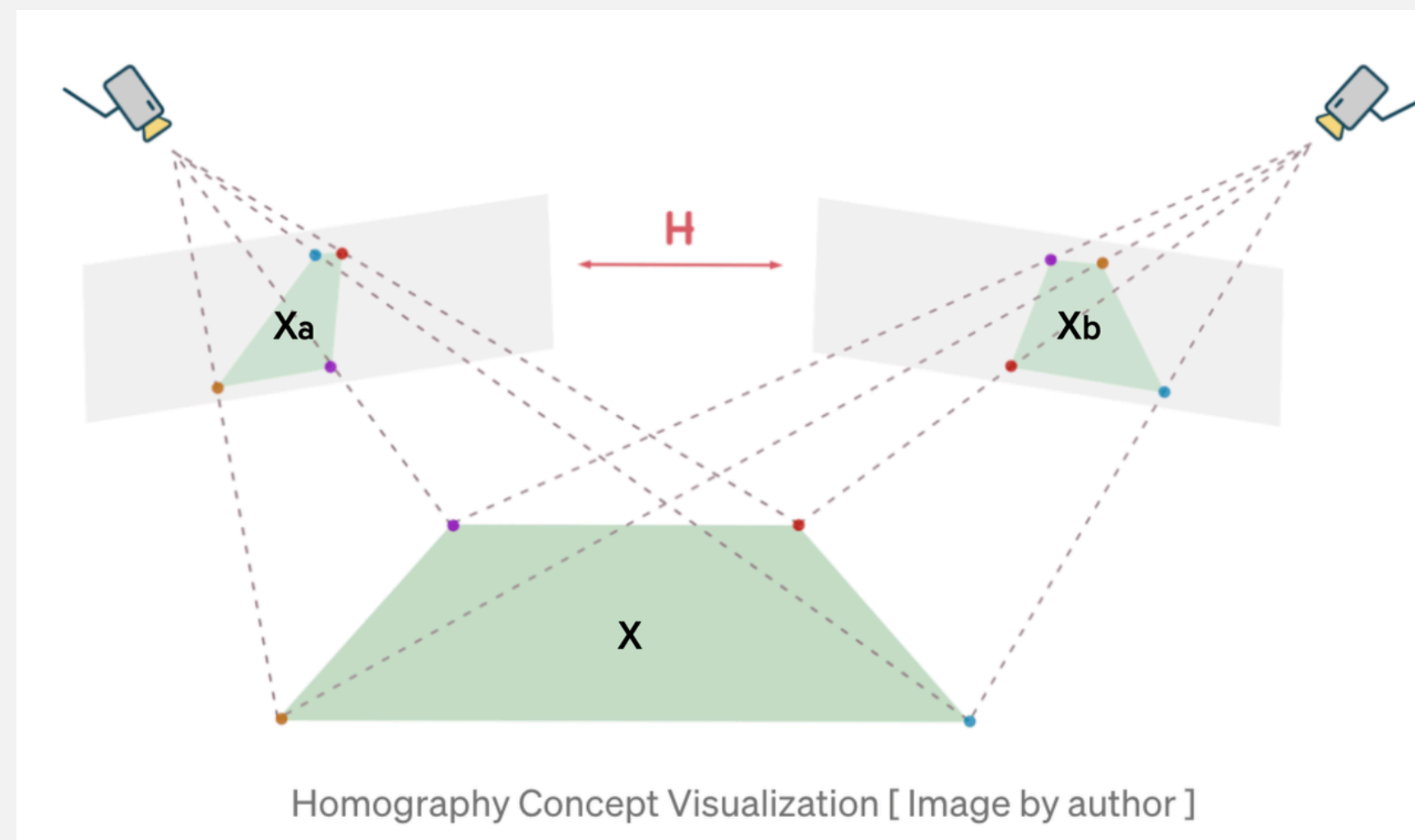


Encontrando um estudante em mais de uma câmera

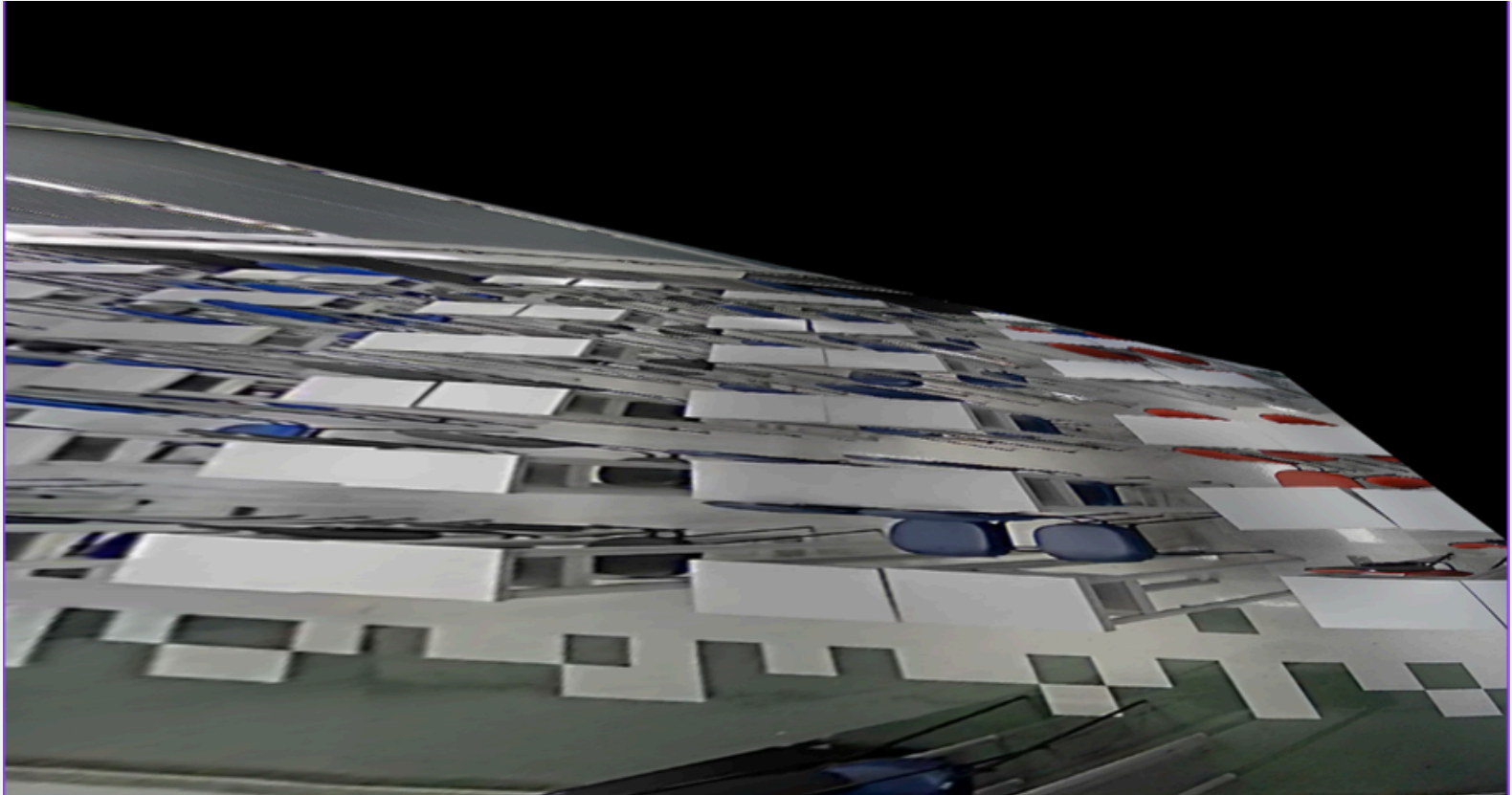


Processo de Homografia

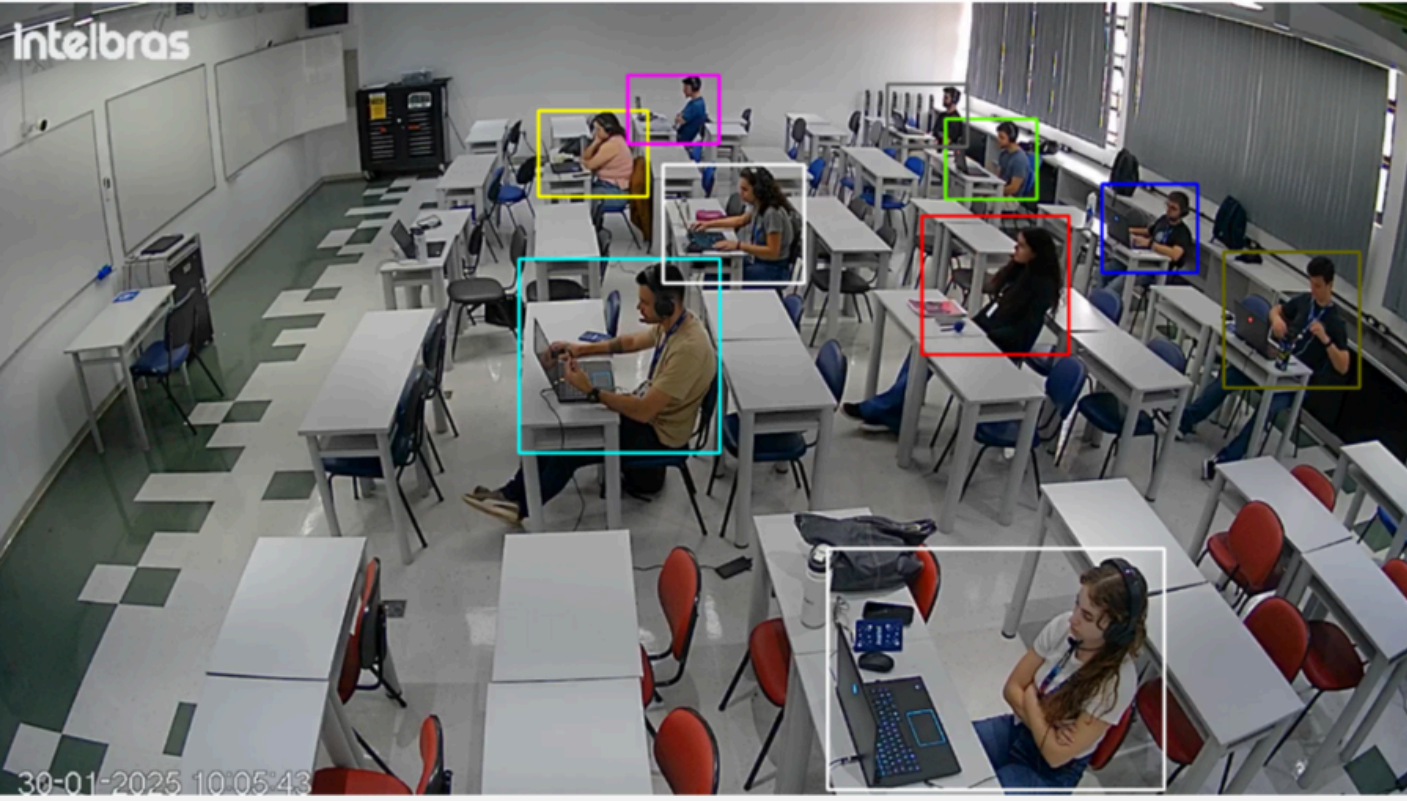
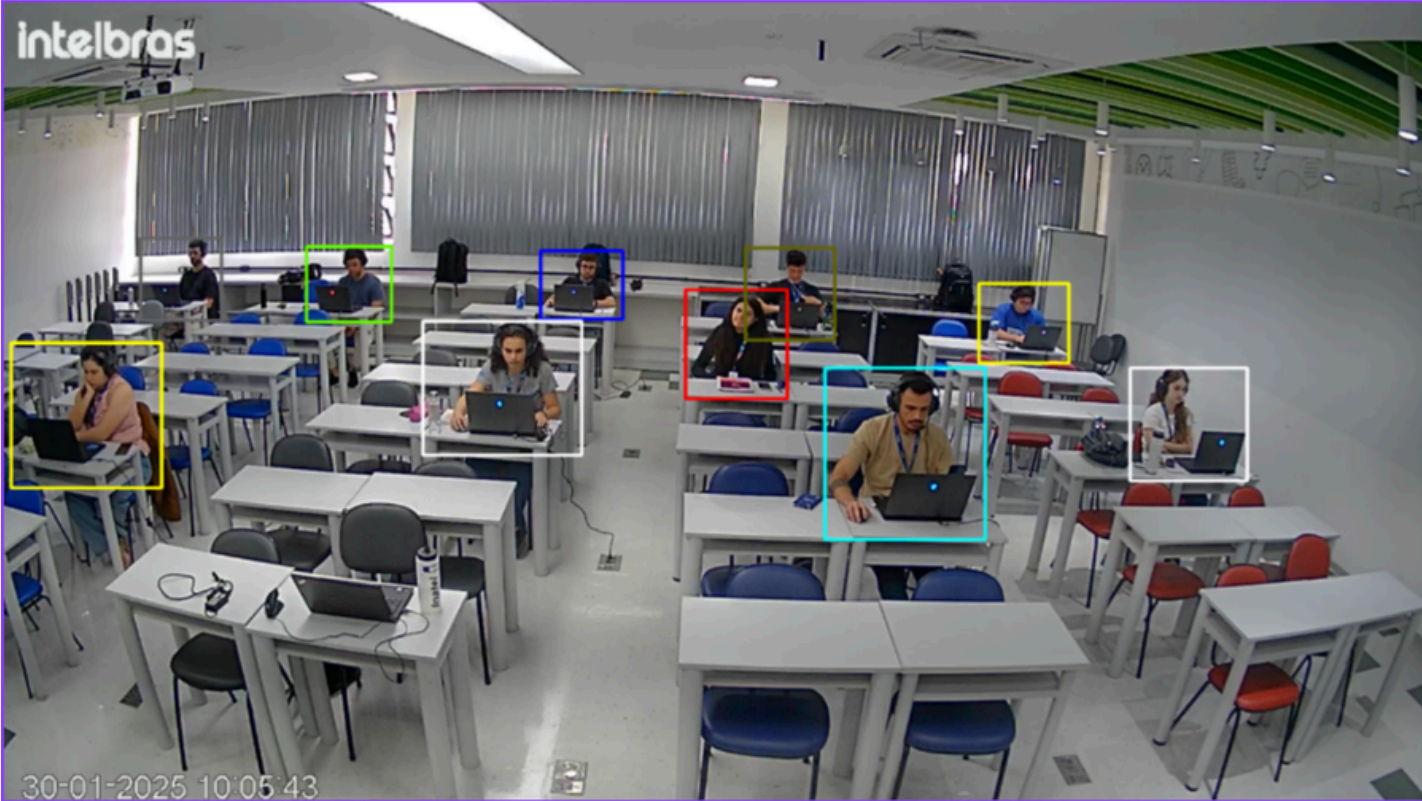
Em computação visual, uma homografia é uma transformação projetiva que relaciona dois planos projetivos, ou seja, mapeia pontos de uma imagem para outra



Processo de Homografia

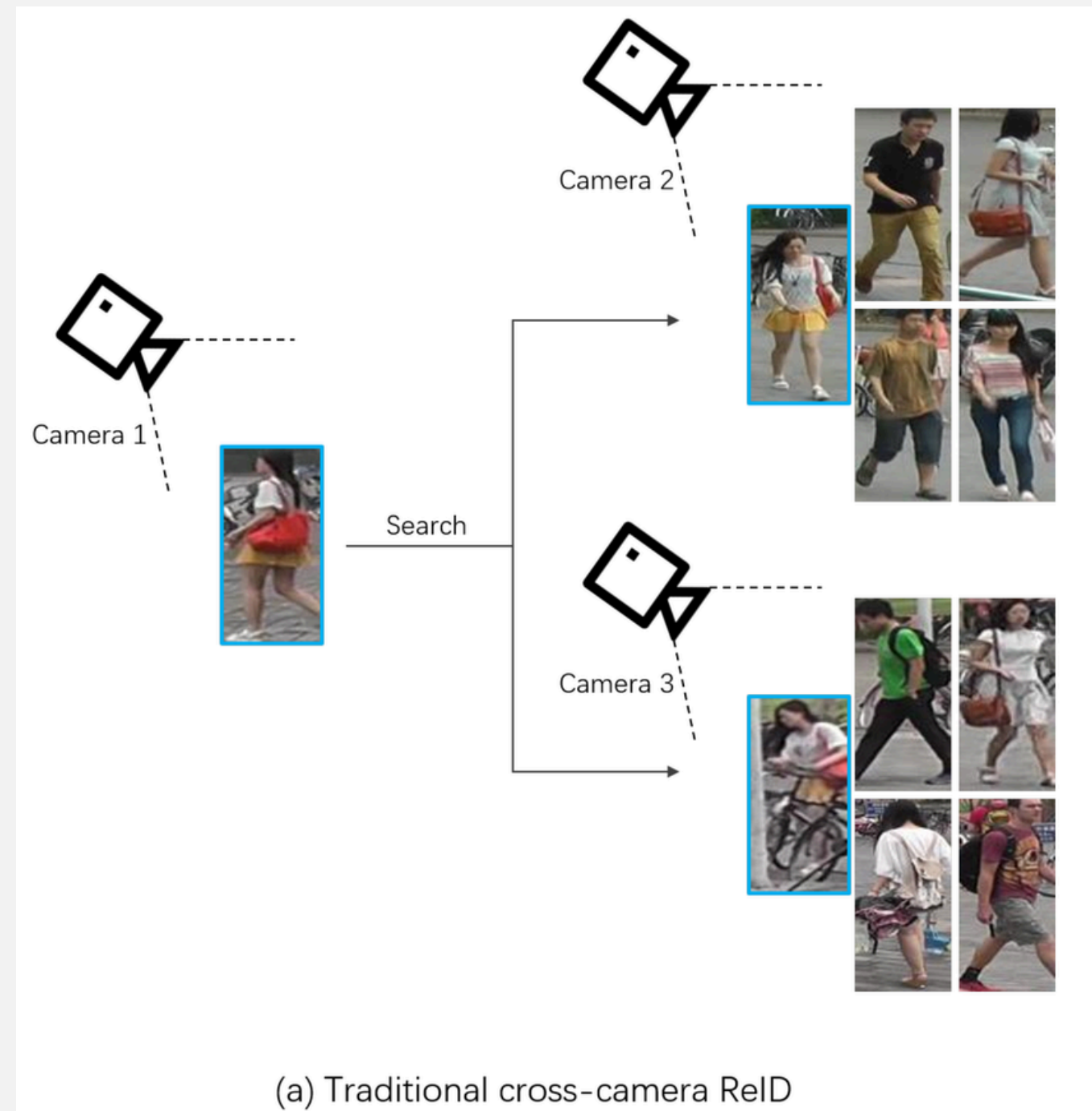


Resultado da Homografia em Sala de Aula



Person REID

Person REID é uma tecnologia de visão computacional que identifica e rastreia a mesma pessoa em diferentes câmeras, mesmo com mudanças de ângulo, iluminação ou roupas



TorchReid

Torchreid é uma biblioteca para reidentificação de pessoas por meio de aprendizado profundo, escrita em PyTorch.

ImageNet classification models

- ResNet
- ResNeXt
- SENet
- DenseNet
- Inception-ResNet-V2
- Inception-V4
- Xception
- IBN-Net

Lightweight models

- NASNet
- MobileNetV2
- ShuffleNet
- ShuffleNetV2
- SqueezeNet

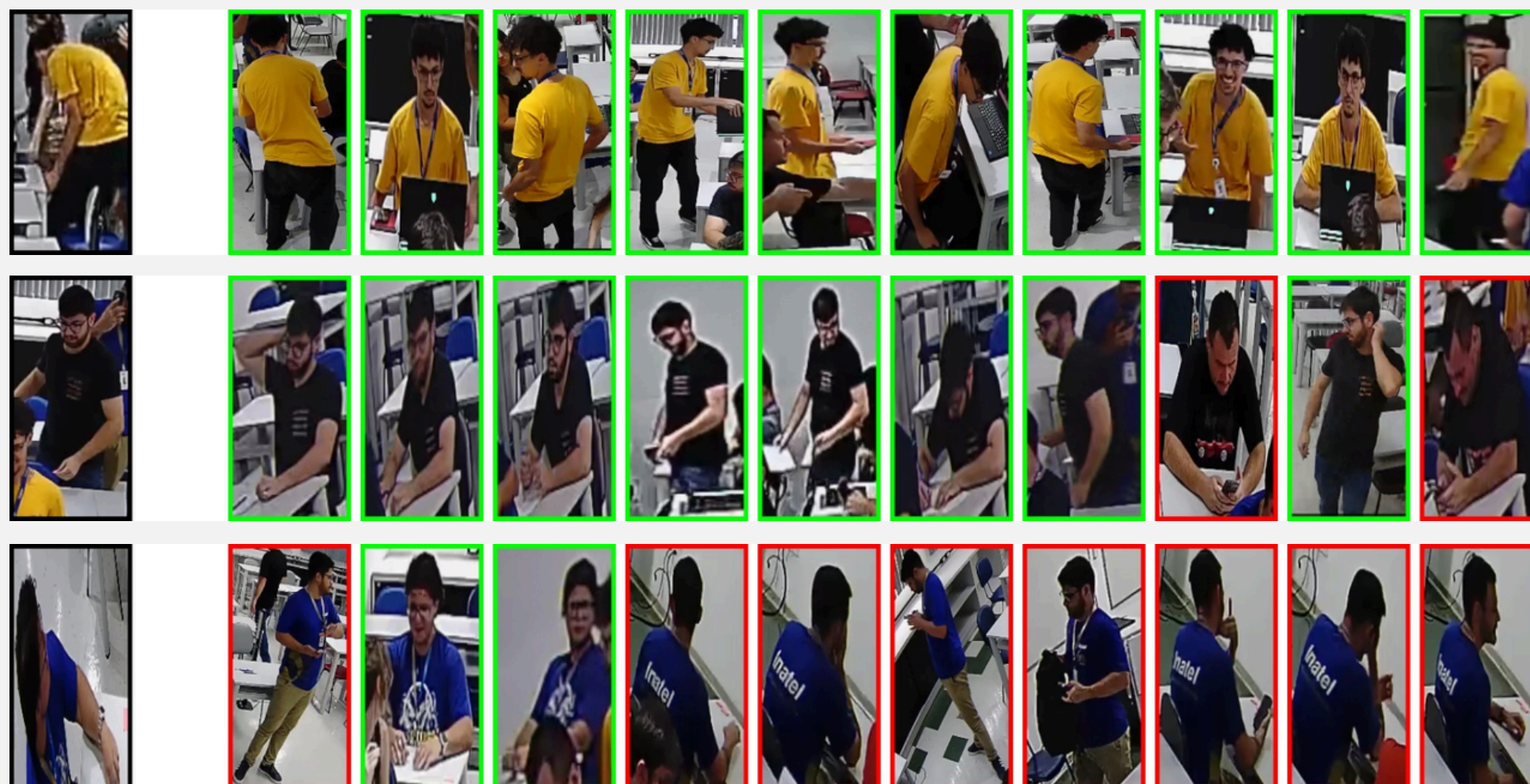
ReID-specific models

- MuDeep
- ResNet-mid
- HACNN
- PCB
- MLFN
- OSNet
- OSNet-AIN

Resultados Iniciais

Resultados iniciais para a rede osnet_x1_0

```
##### Evaluating data_set (source) #####  
Extracting features from query set ...  
Done, obtained 712-by-512 matrix  
Extracting features from gallery set ...  
Done, obtained 712-by-512 matrix  
Speed: 0.0764 sec/batch  
Computing distance matrix with metric=euclidean ...  
Computing CMC and mAP ...  
** Results **  
mAP: 51.8%  
CMC curve  
Rank-1 : 69.9%  
Rank-5 : 89.7%  
Rank-10 : 95.2%  
Rank-20 : 98.3%
```

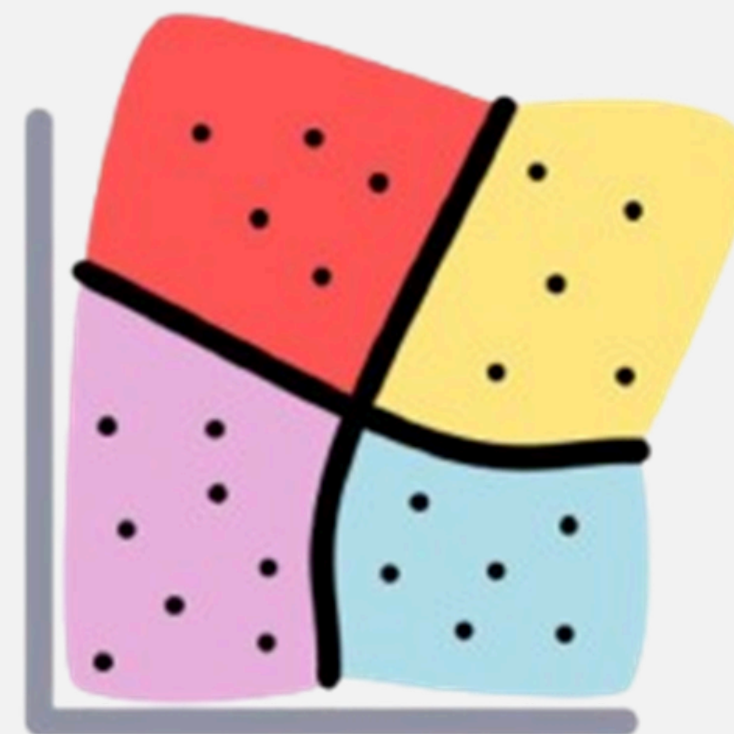


Perfil de engajamento observacional

As estimativas de engajamento observacional são criadas a partir de combinações de ações. Para o perfil de engajamento será observado o conjunto e combinações das ações no tempo transcorrido de aula.

Nível de engajamento

- Alto
- Médio
- Baixo
- Desengajado
- Ações

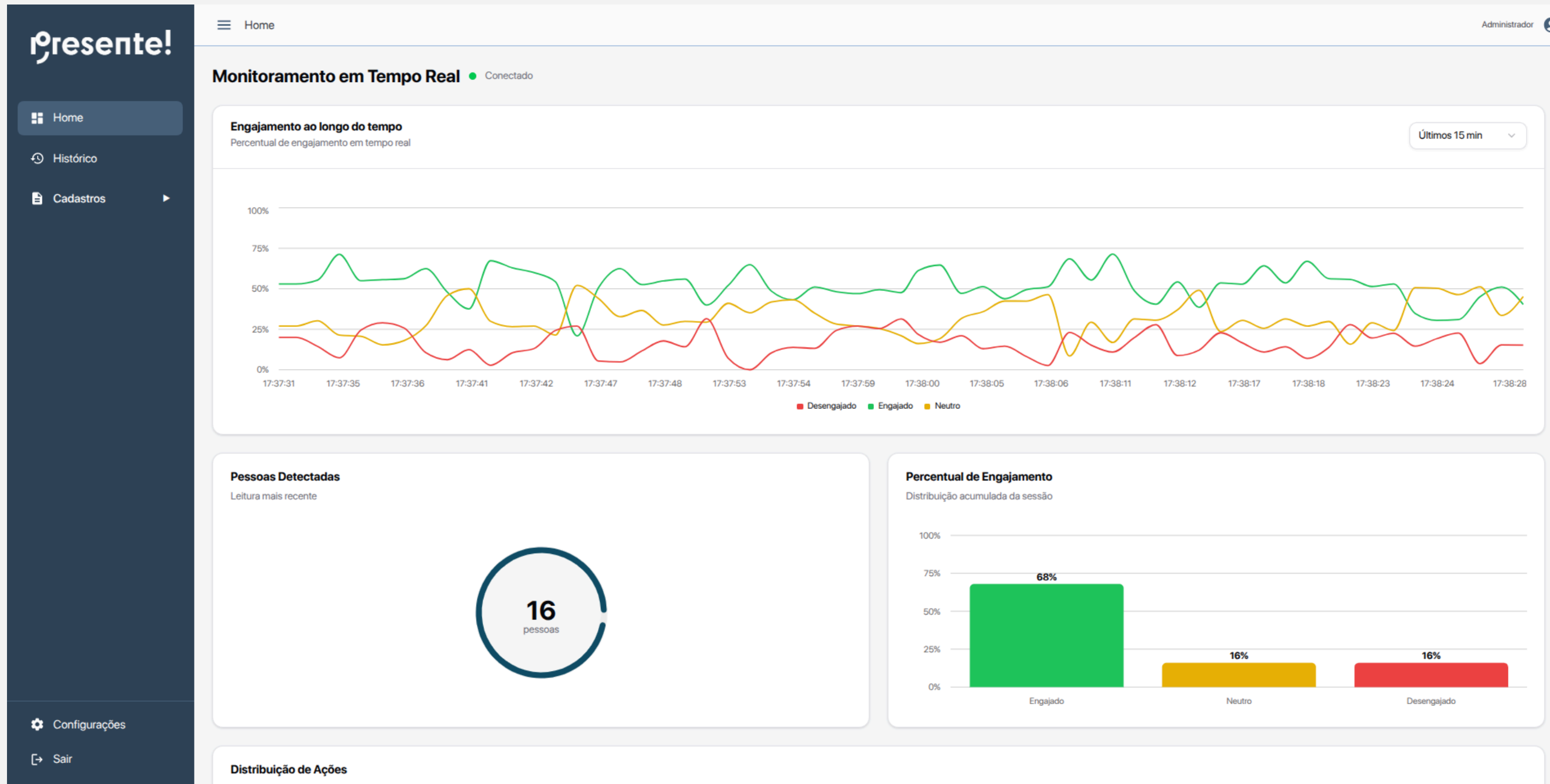


Clusterização de ações

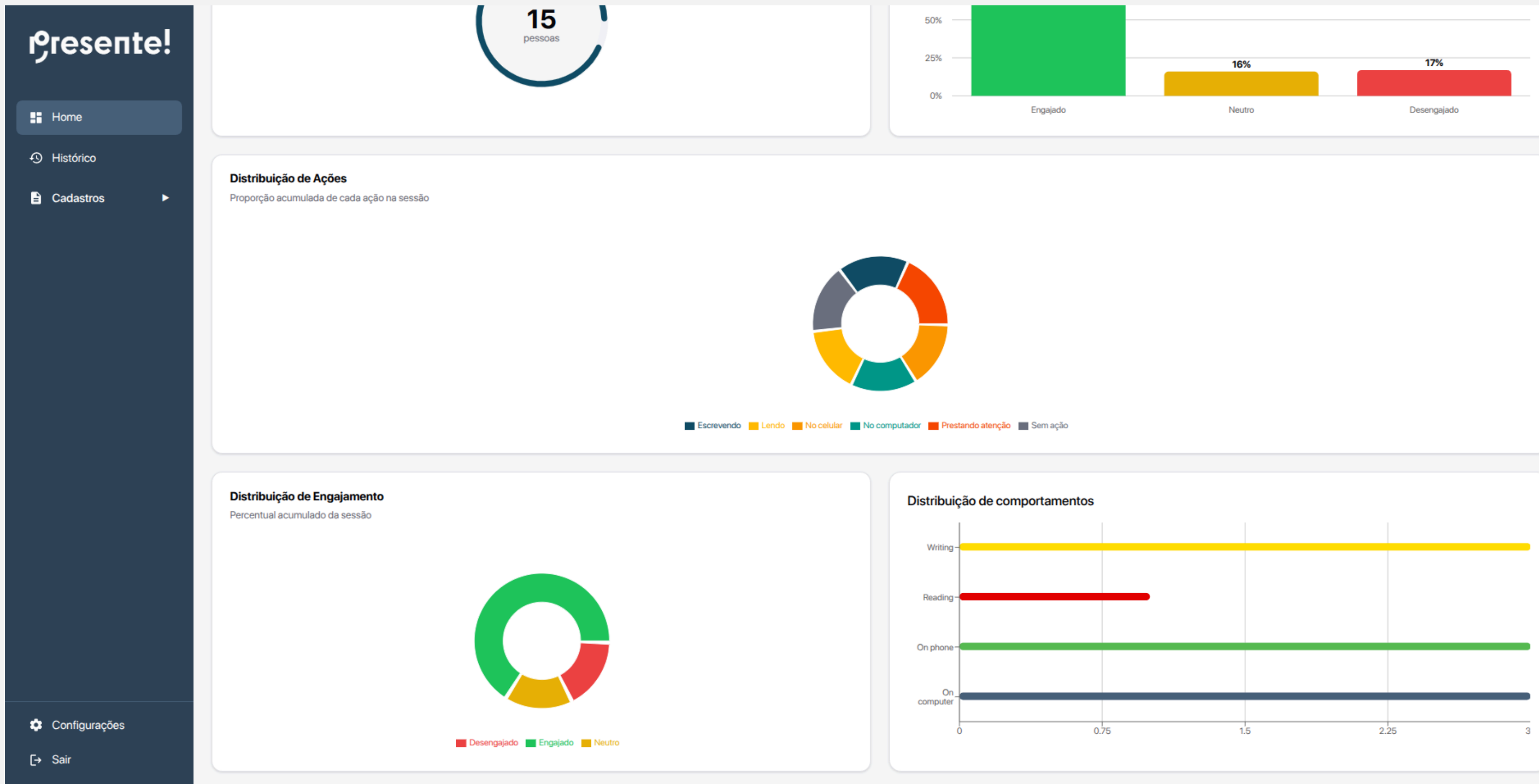


Perfil de engajamento

Perfil de engajamento observacional



Perfil de engajamento observacional



Obrigado!

 **presente!**

Sua sala de aula
inteligente.