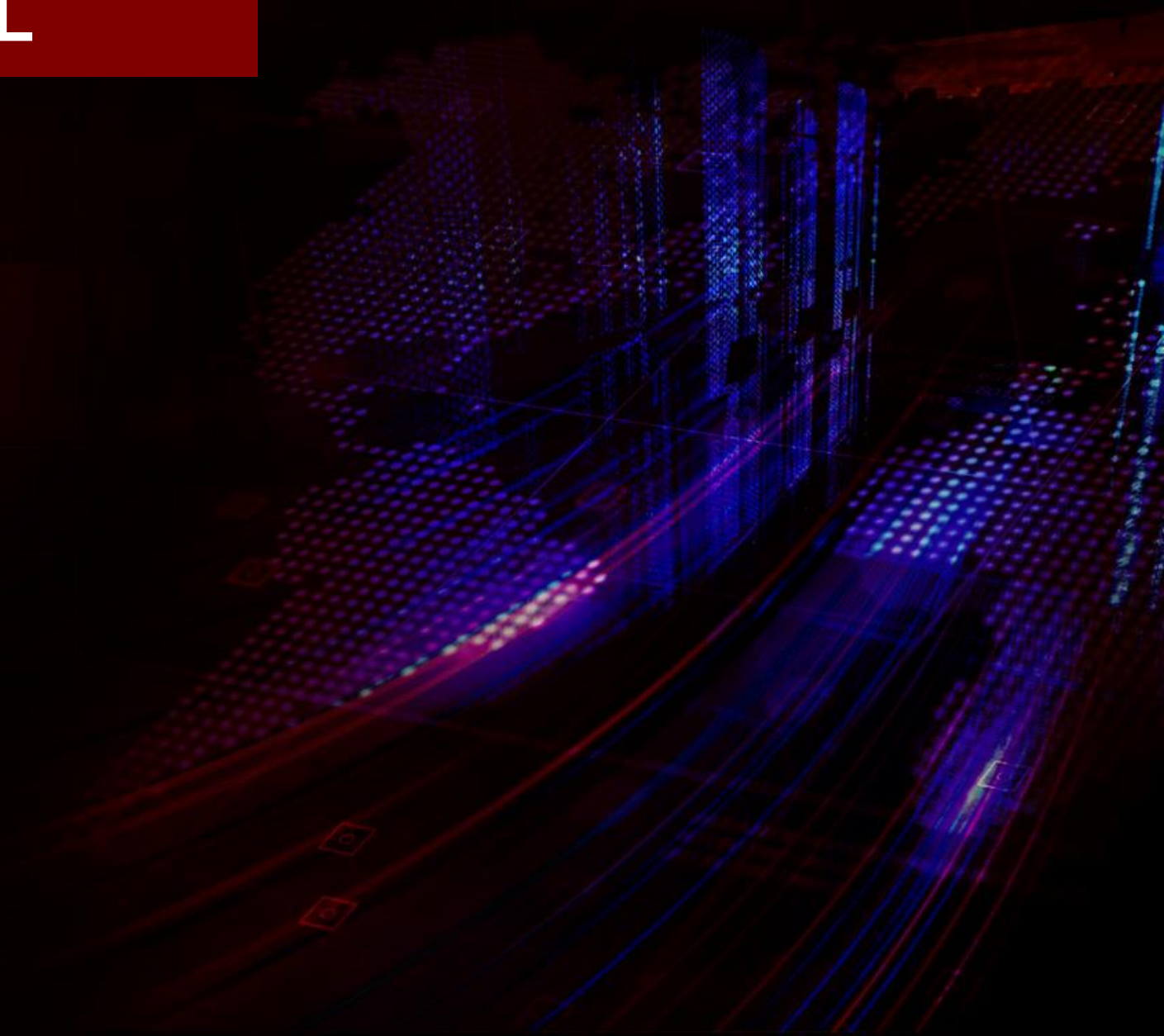


VISÃO COMPUTACIONAL

Reconhecimento de objetos



TÓPICOS

1. Aplicações
2. Detecção de objetos
3. Segmentação de instância



APLICAÇÕES

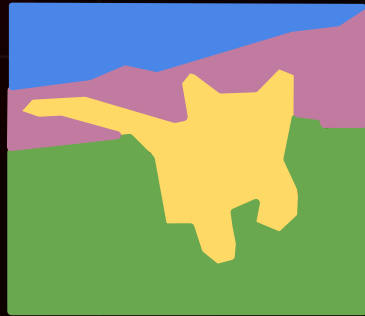
Classificação



GATO

Sem
extensão
espacial

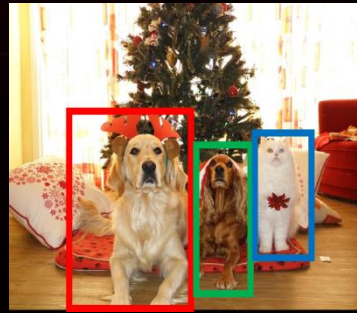
Segmentação semântica



GRAMA
GATO
ÁRVORE CÉU

Nenhum objeto,
apenas pixels

Detecção de objetos



CÃO CÃO
GATO

Múltiplos objetos

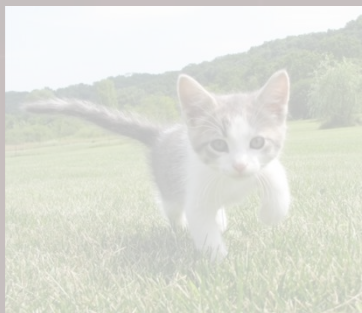
Segmentação de instância



CÃO CÃO
GATO

DETECÇÃO DE OBJETOS

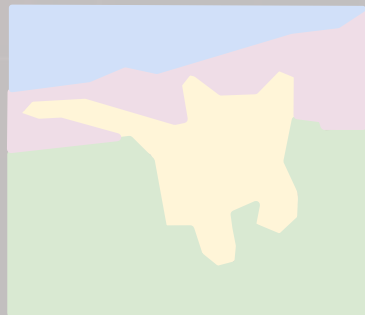
Classificação



GATO

Sem
extensão
espacial

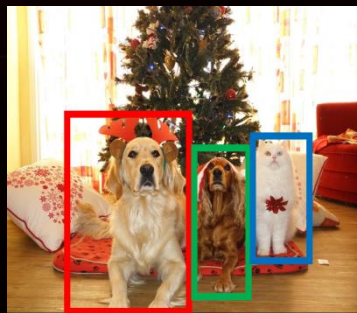
Segmentação semântica



GRAMA
GATO
ÁRVORE CÉU

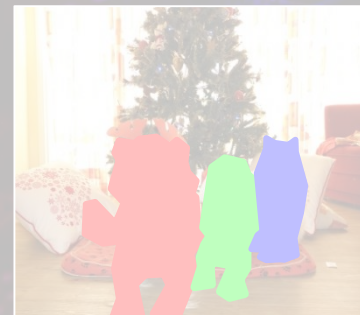
Nenhum objeto,
apenas pixels

Deteccção de objetos



CÃO CÃO
GATO

Segmentação de instância



CÃO CÃO
GATO

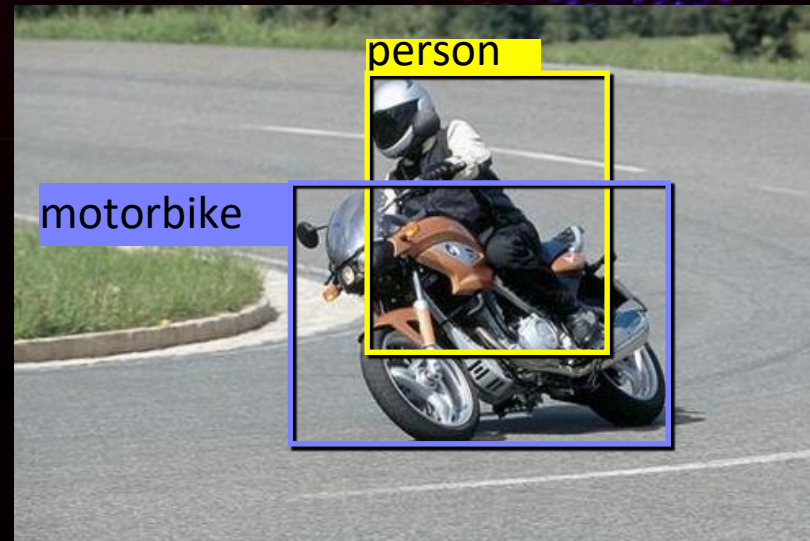
Múltiplos objetos

FORMULAÇÃO DO PROBLEMA

{ avião, pássaro, moto, pessoa, sofá }



Entrada



Saída desejada

FORMULAÇÃO DO PROBLEMA

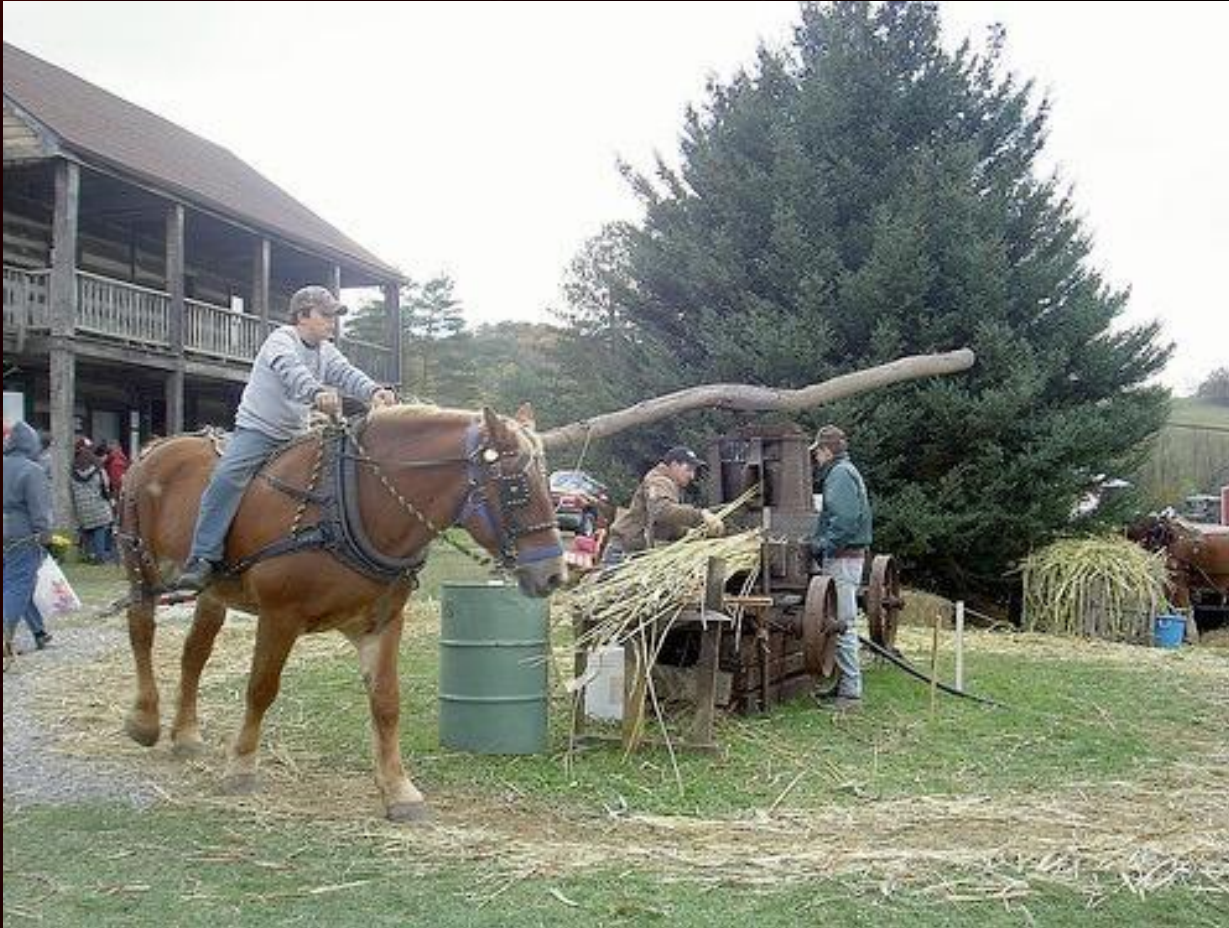
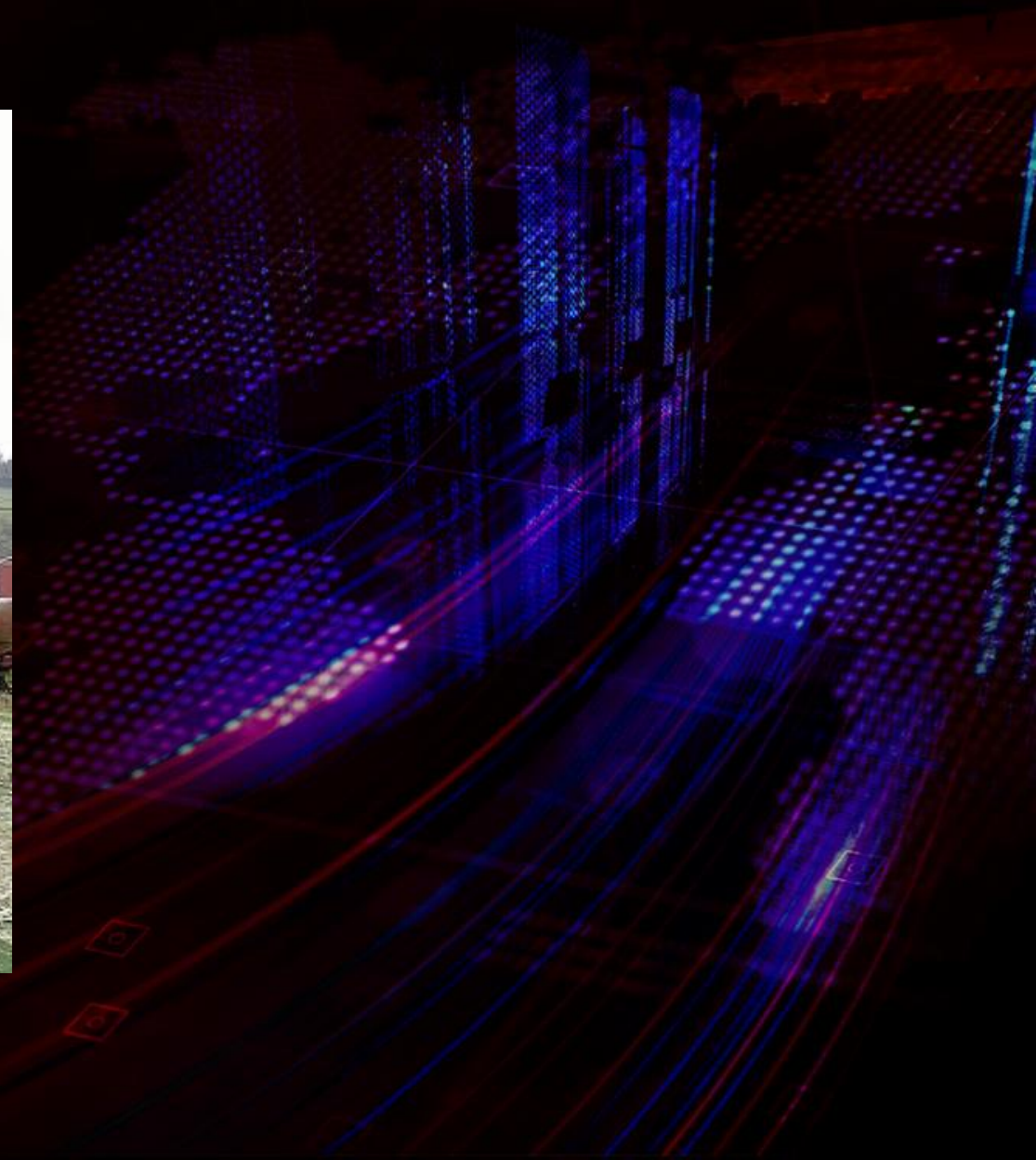


Imagem de teste (não vista)

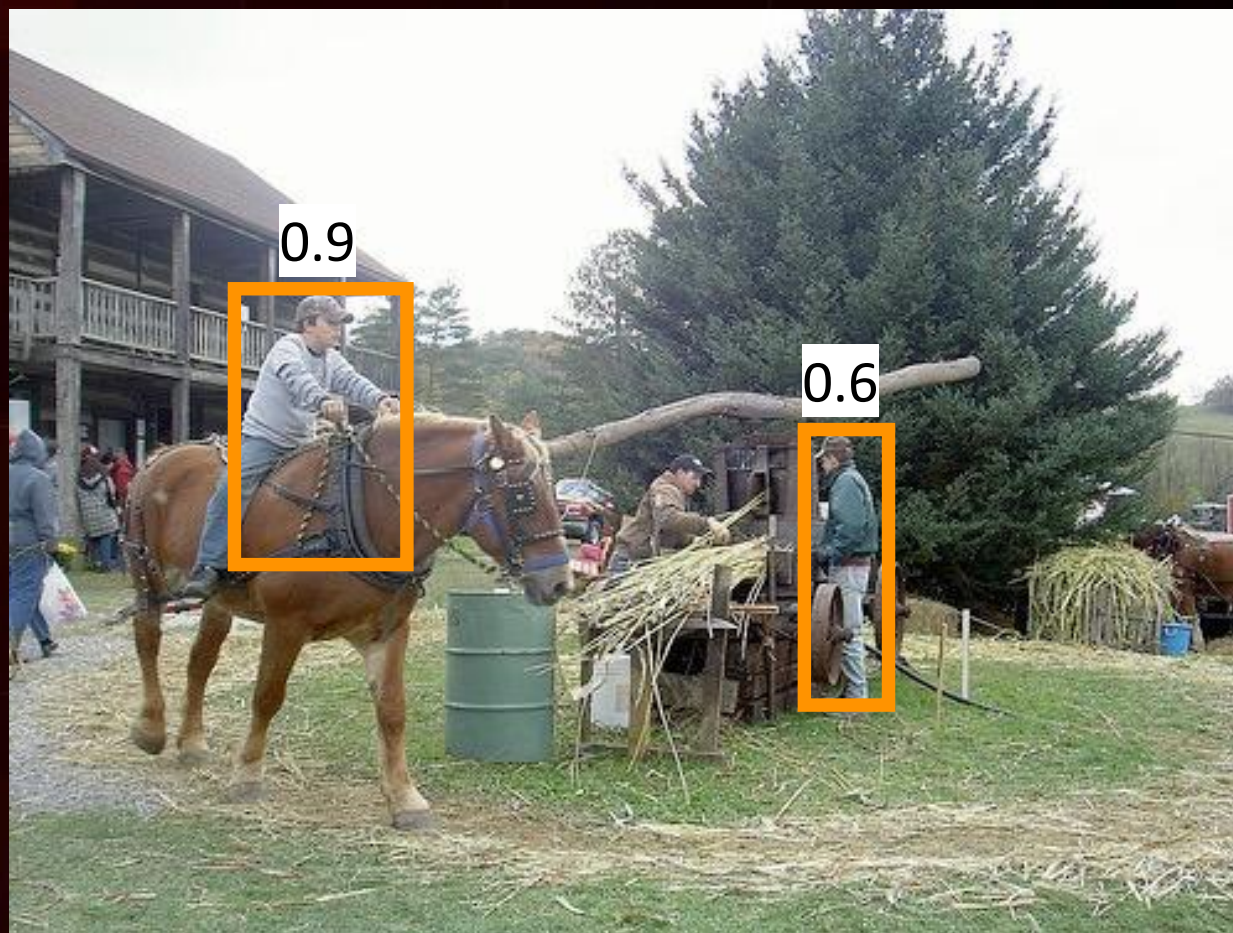
PRIMEIRA DETECÇÃO




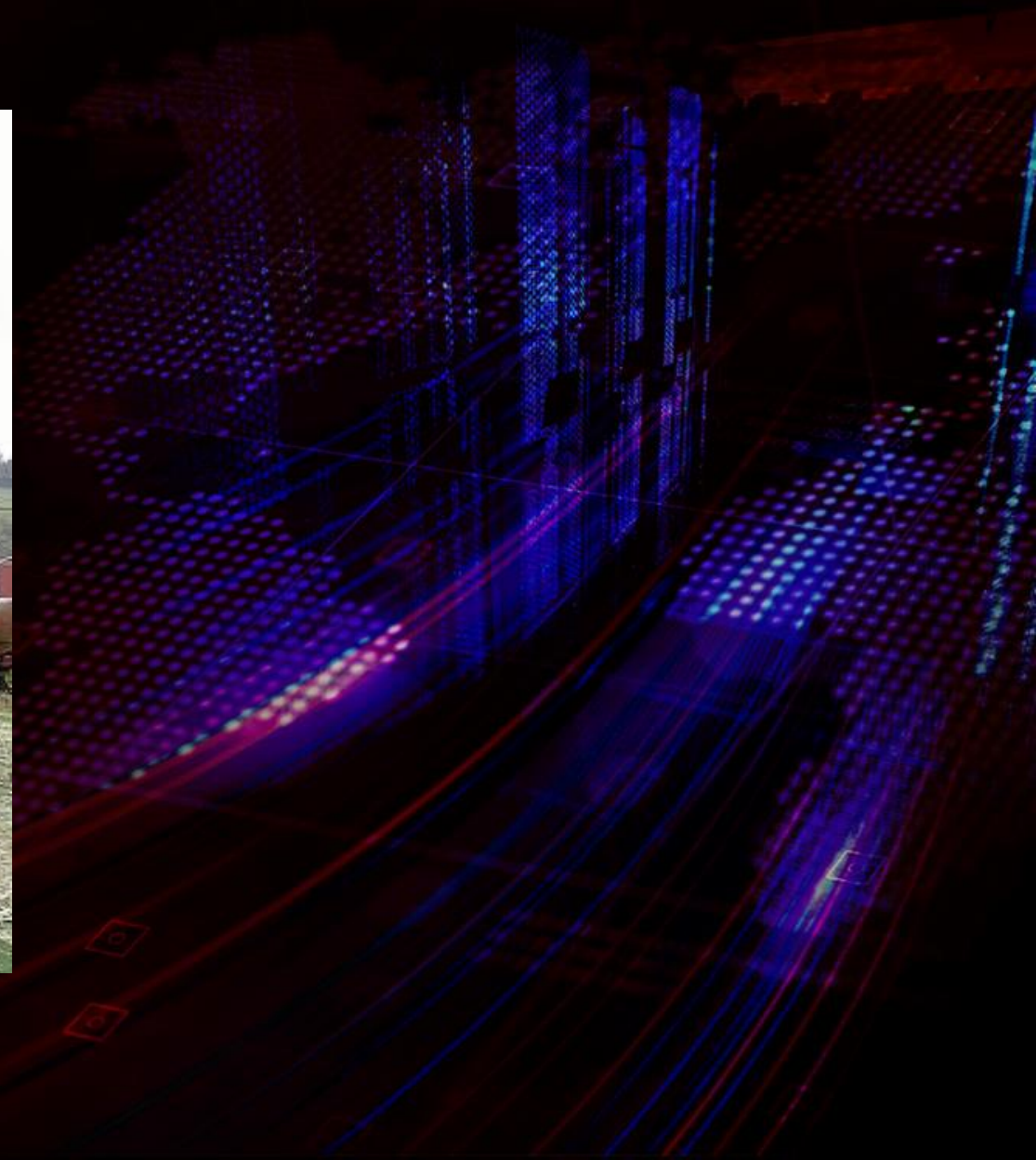
 predições



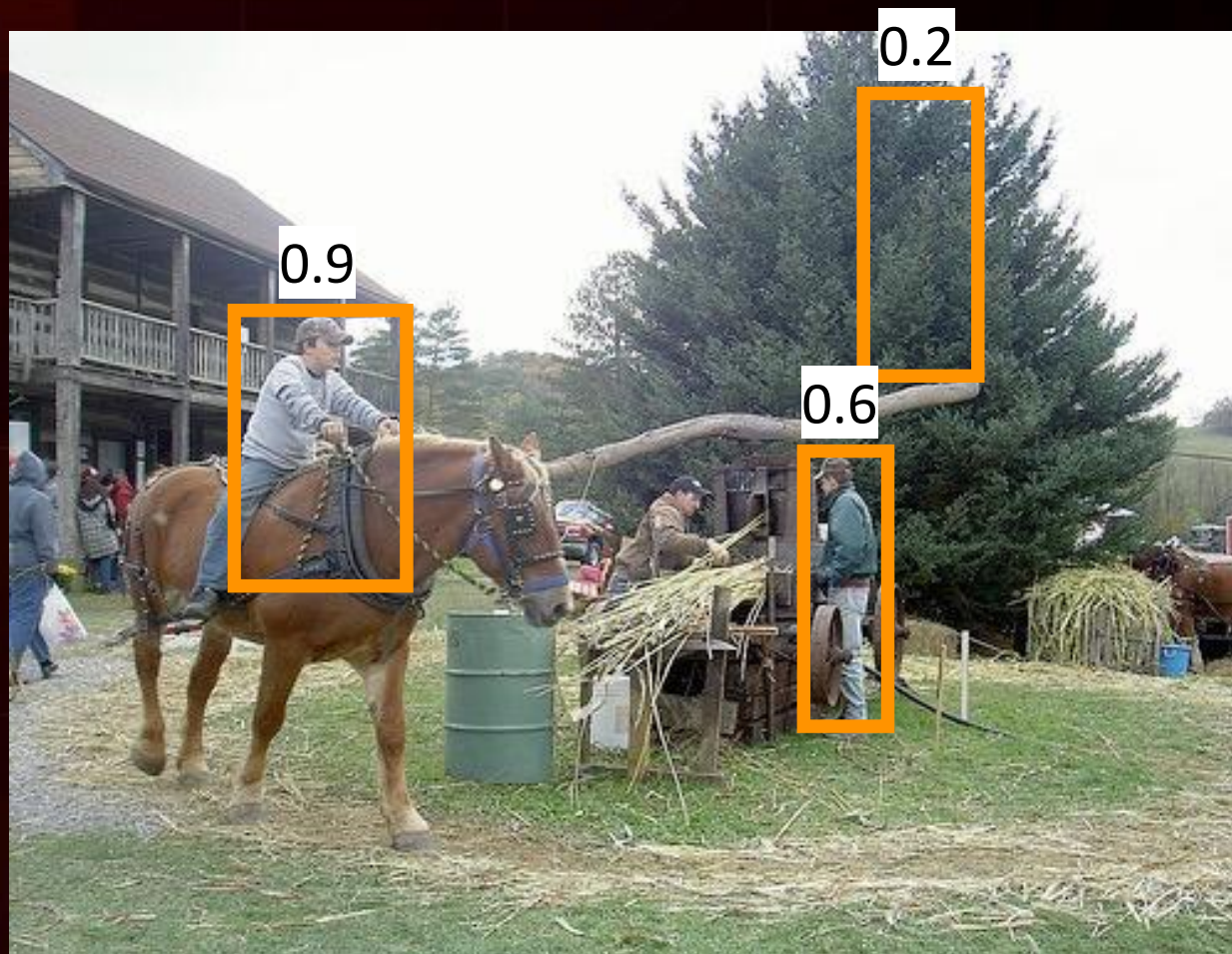
SEGUNDA DETECÇÃO




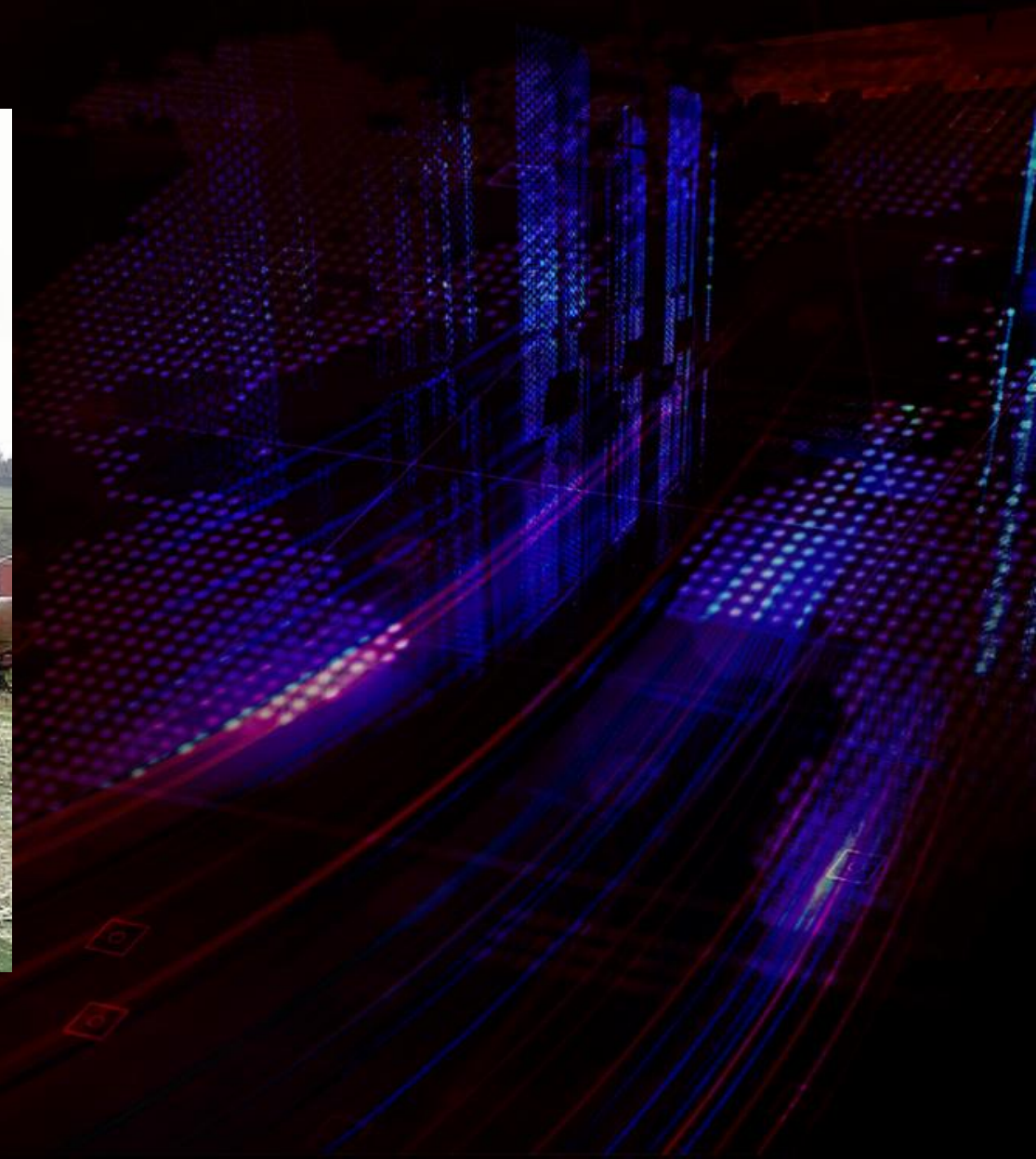
 predições



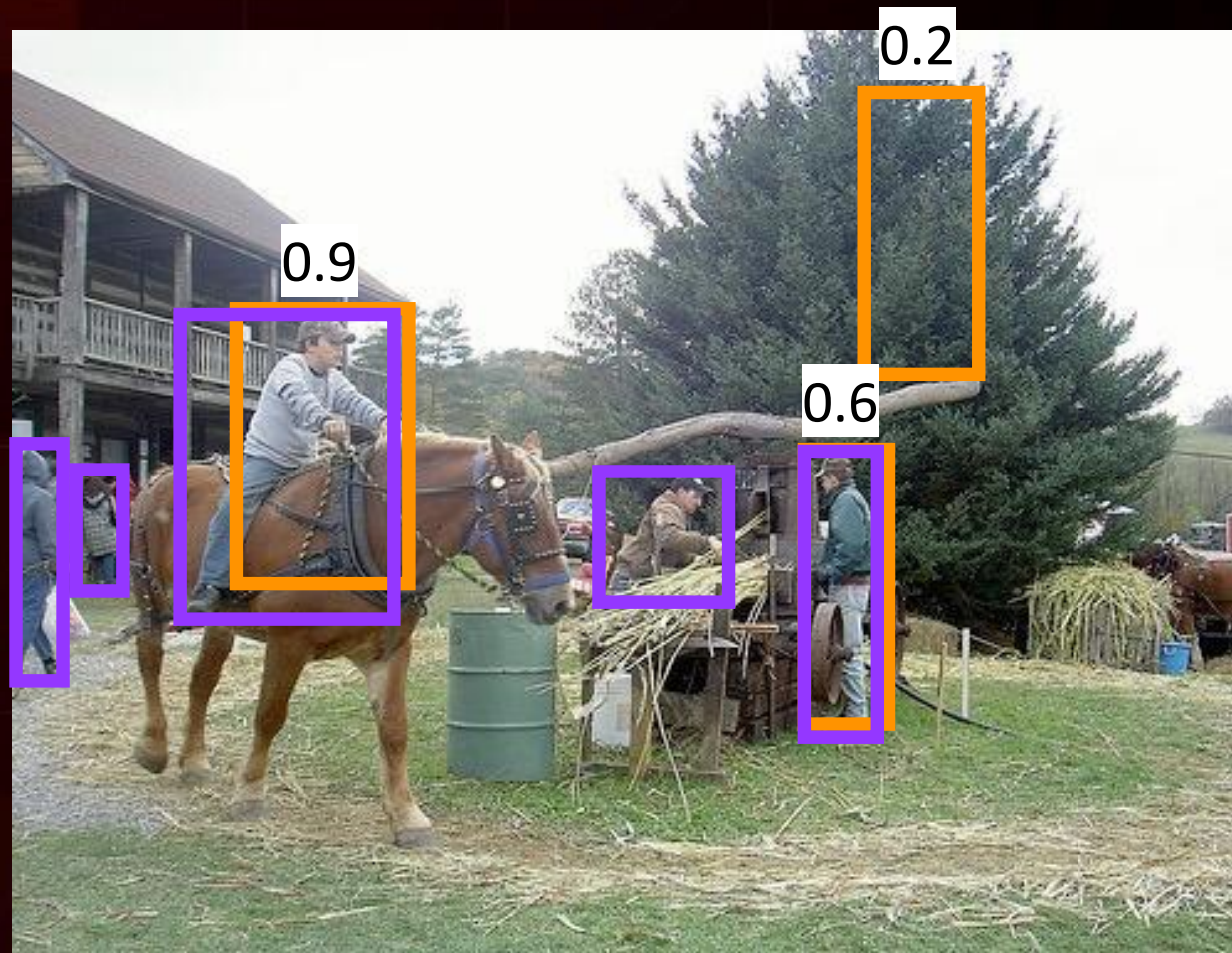
TERCEIRA DETECÇÃO



 predições



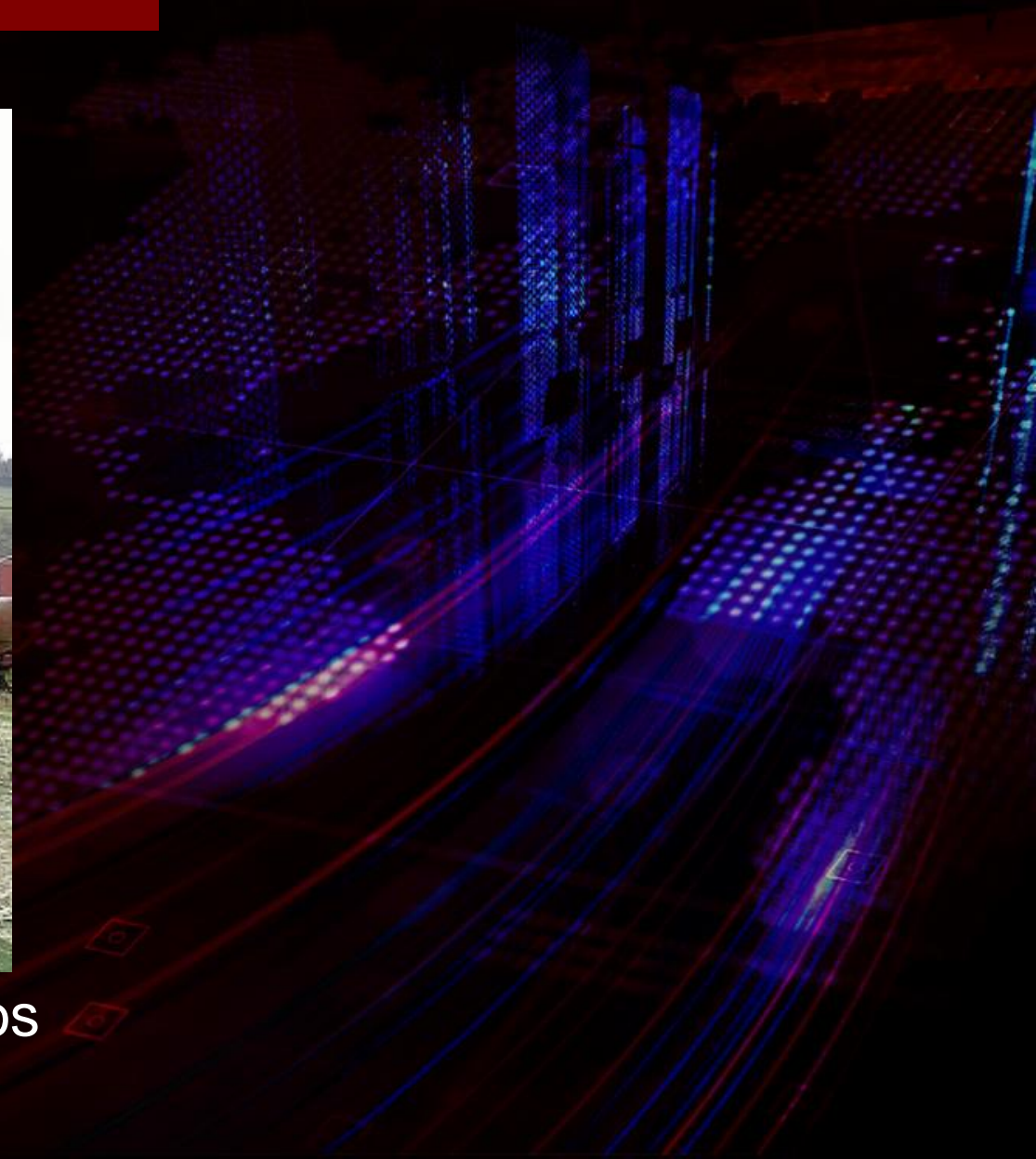
COMPARAÇÃO COM DESEJADO



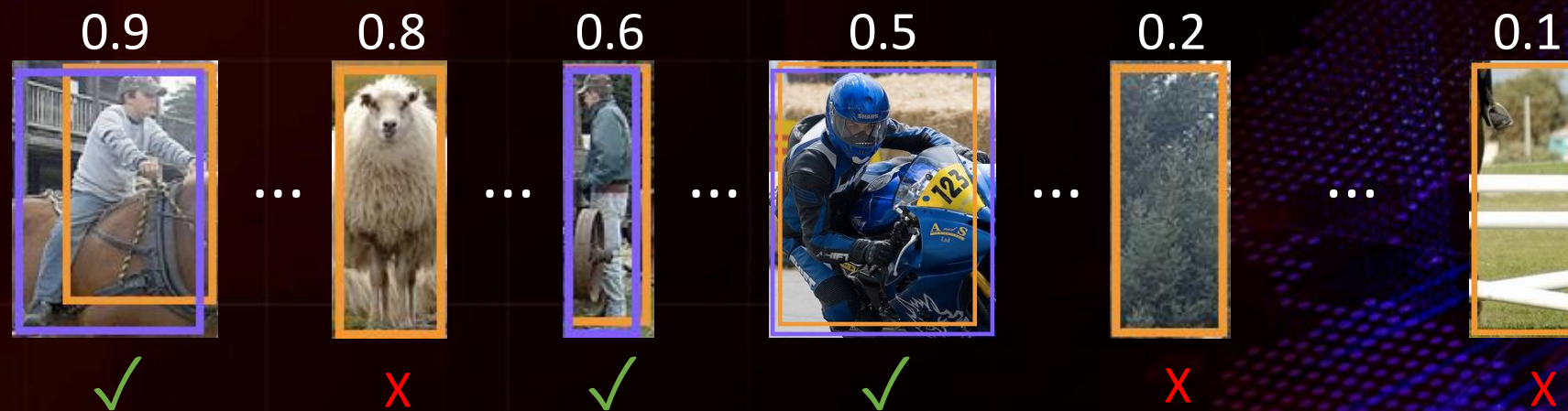
predições



valores desejados



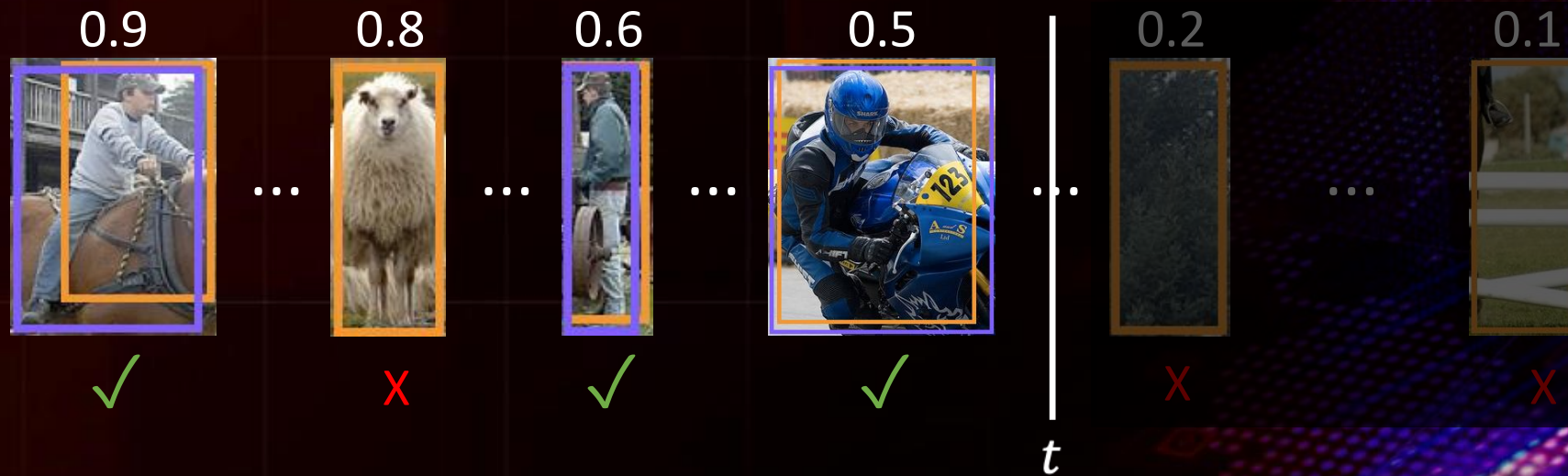
ORDENAR POR CONFIANÇA



**Verdadeiro
Positivo (VP)**
(alta
sobreposição)

**Falso
Positivo (FP)**
(sem sobreposição,
baixa sobreposição,
ou duplicado)

MÉTRICA DE AVALIAÇÃO

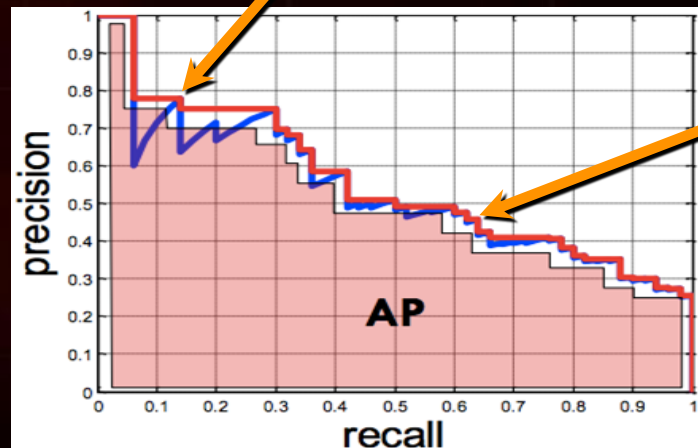
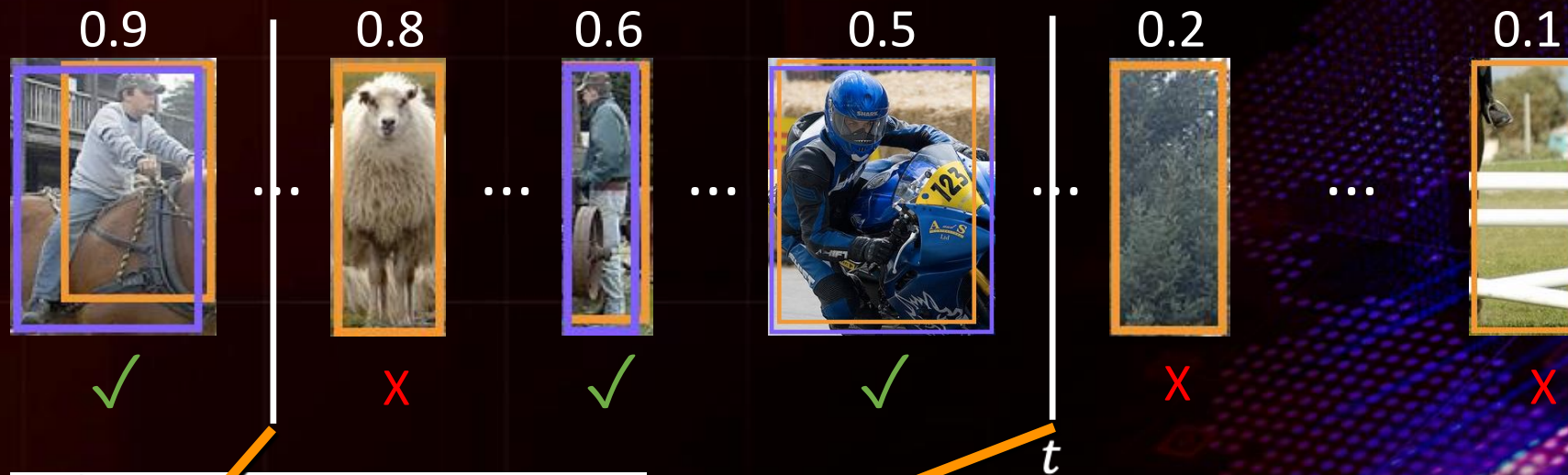


$$precisão@t = \frac{\#VP@t}{\#VP@t + \#FP@t}$$

$\frac{\text{✓}}{\text{✓} + \text{✗}}$

$$recall@t = \frac{\#VP@t}{\#VP@t + \#FN@t} \longrightarrow \# \text{ total objetos positivos}$$

MÉTRICA DE AVALIAÇÃO



Average Precision (AP)

0% é pior

100% é melhor

mean AP sobre classes (mAP)

MÉTRICA DE AVALIAÇÃO

- Como são tomadas as decisões (detecções) para o cálculo das métricas de avaliação?
 - Casamento (*Matching*)
 - Distância
 - Correlação
- Aprendizado de Máquina Supervisionado
 - Classificadores
 - Bayesianos
 - SVMs (*Support Vector Machines*)
 - Redes Neurais Artificiais

EXEMPLO: DETECÇÃO DE PESSOA

Dalal & Triggs '05:

Treinar **SVM** usando descritores **HOG**

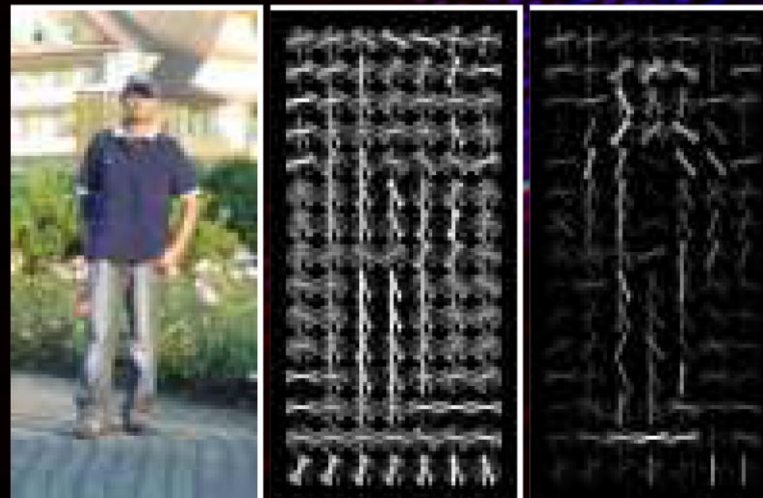
2 classes, pessoa / não pessoa

No teste:

Extrair HOGs em várias escalas

Rodar SVM em todas localizações

Respostas = Pessoa?



EXEMPLO: DETECÇÃO DE PESSOA

Detector de janela deslizante

Muitas escalas

Vários locais



EXEMPLO: DETECÇÃO DE PESSOA

Detector de janela deslizante

Muitas escalas

Vários locais

Pessoa? Não



EXEMPLO: DETECÇÃO DE PESSOA

Detector de janela deslizante

Muitas escalas

Vários locais

Pessoa? Não



EXEMPLO: DETECÇÃO DE PESSOA

Detector de janela deslizante

Muitas escalas

Vários locais

Pessoa? Não



EXEMPLO: DETECÇÃO DE PESSOA

Detector de janela deslizante

Muitas escalas

Vários locais

Pessoa? Não



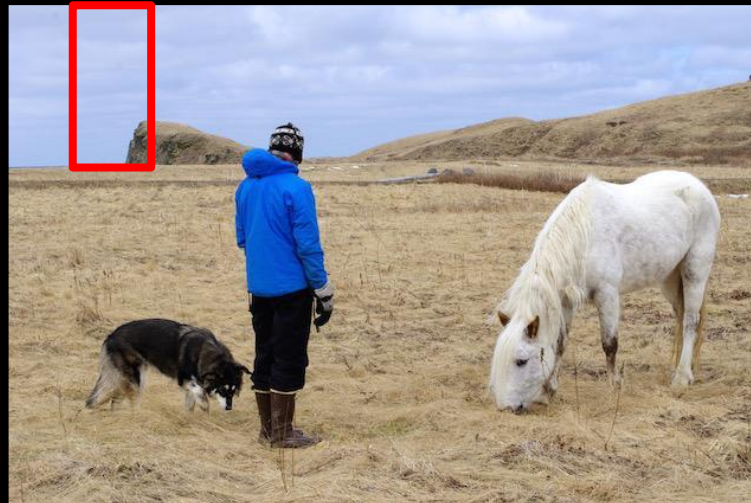
EXEMPLO: DETECÇÃO DE PESSOA

Detector de janela deslizante

Muitas escalas

Vários locais

Pessoa? Não



EXEMPLO: DETECÇÃO DE PESSOA

Detector de janela deslizante

Muitas escalas

Vários locais



QUESTÕES...

Aprendizado de Máquina precisa de características!

Quais são os melhores descritores?

HOG?

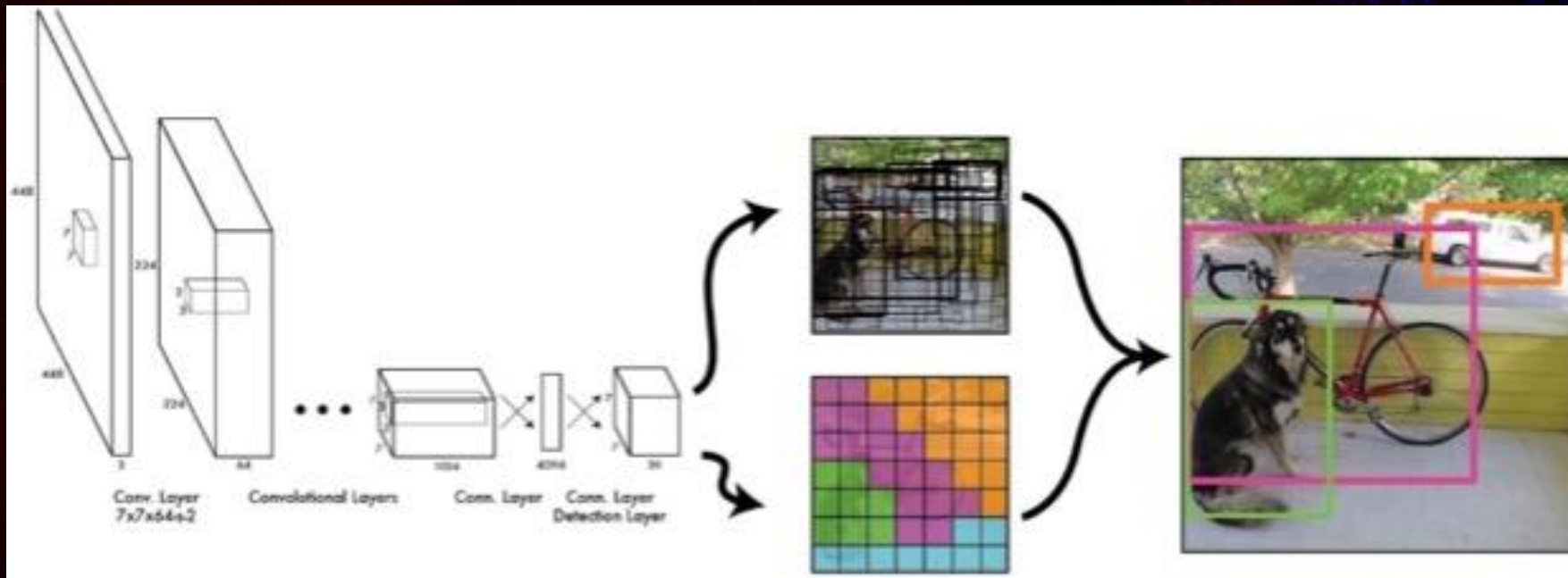
SIFT?

Por que não deixar o algoritmo decidir?

**Redes Neurais: extração de características +
modelo linear**

SUCESSO DAS REDES NEURAIS

- É possível treinar uma rede neural (profunda) para realizar o *pipeline* completo



SUCESSO DAS REDES NEURAIS

Classificação de imagens:

54% -> 90% acurácia em 1000 classes

Detecção de objetos:

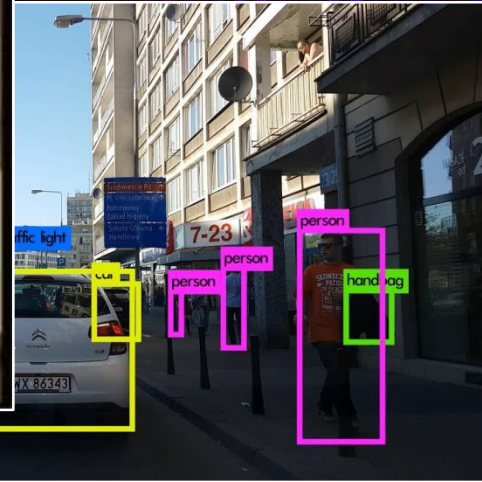
33% mAP (DPM) -> 93% mAP em 20 classes



SEGMENTAÇÃO X DETECÇÃO

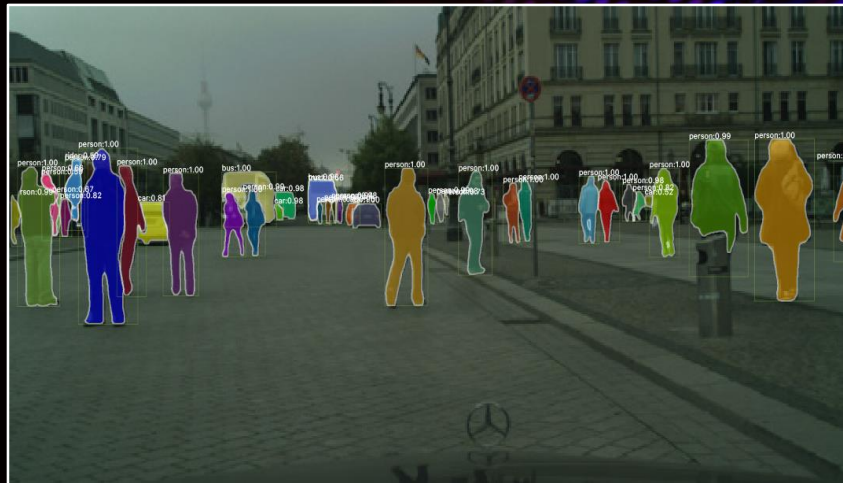
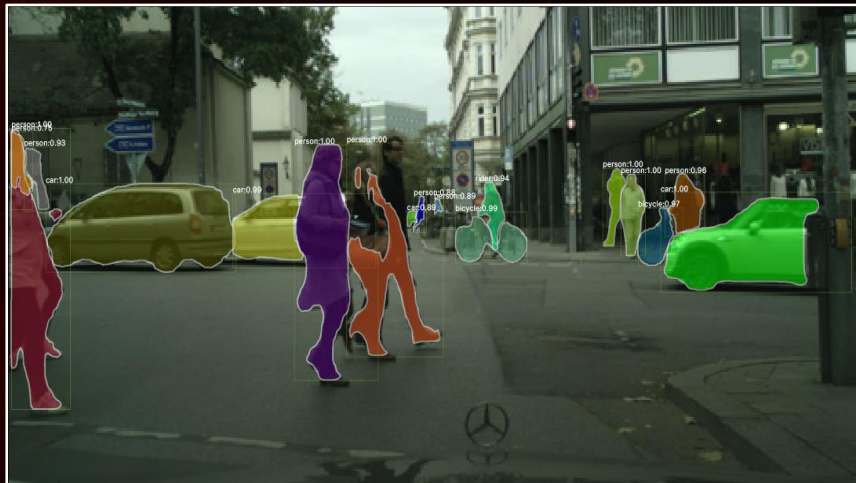
- Rótulos por pixels
- Apenas categoria

Rótulos por 'caixas'
Categoria + instância



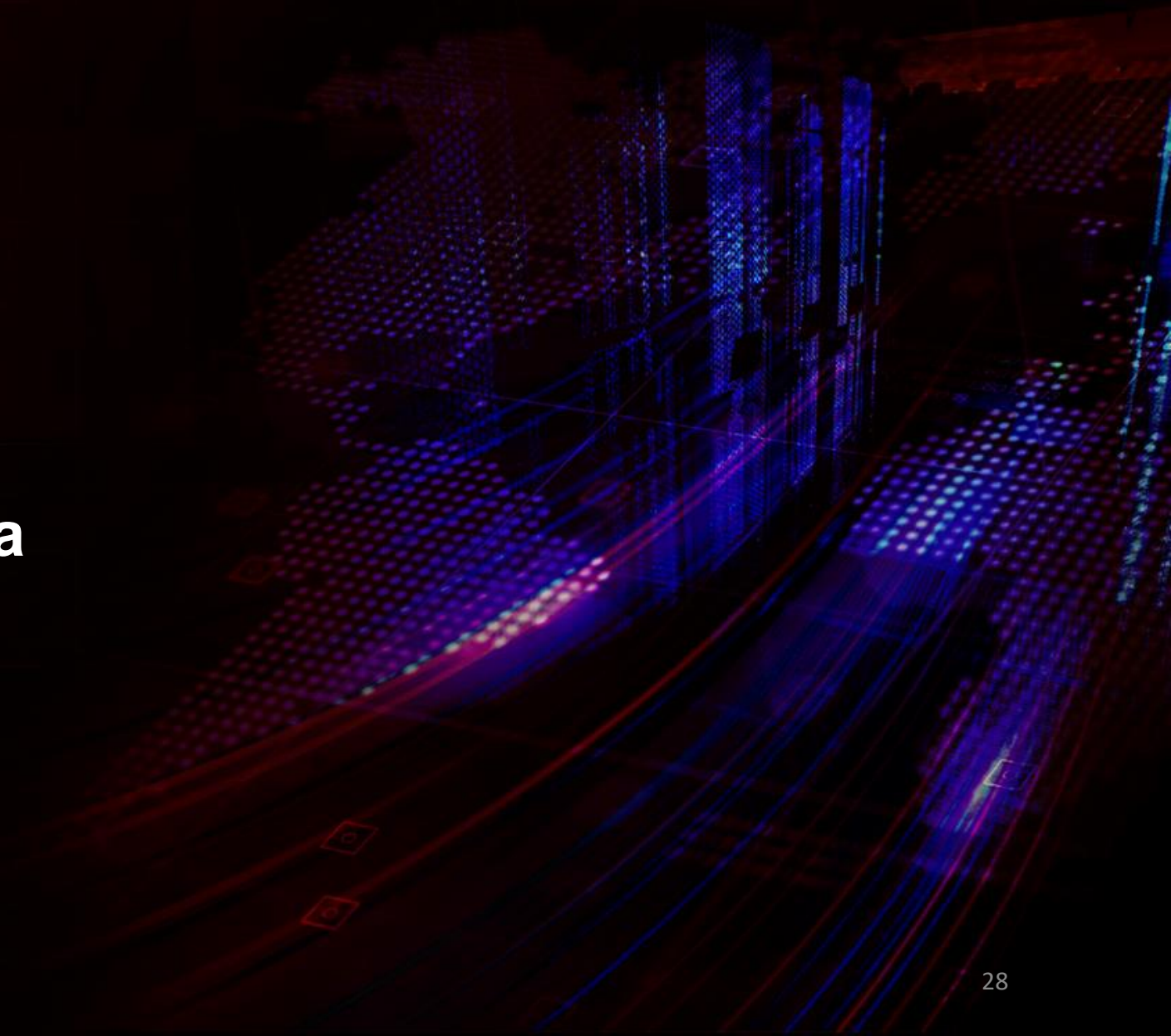
SEGMENTAÇÃO DE INSTÂNCIA

- **Segmentação em nível de instância**
A qual classe cada pixel pertence
Também a qual instância



O QUE VIMOS?

- Aplicações
- Detecção de objetos
- Detecção de pessoa
- Segmentação de instância



PRÓXIMA VIDEOAULA

- Aplicação prática com *Deep Learning*