

Detecção de pessoas em um ambiente industrial utilizando imagens de profundidade e classificadores profundos

Eduardo Henrique Arnold e Danilo Silva

4 de Setembro de 2017

Departamento de Engenharia Elétrica e Eletrônica
Universidade Federal de Santa Catarina

Introdução

Caracterização da aplicação



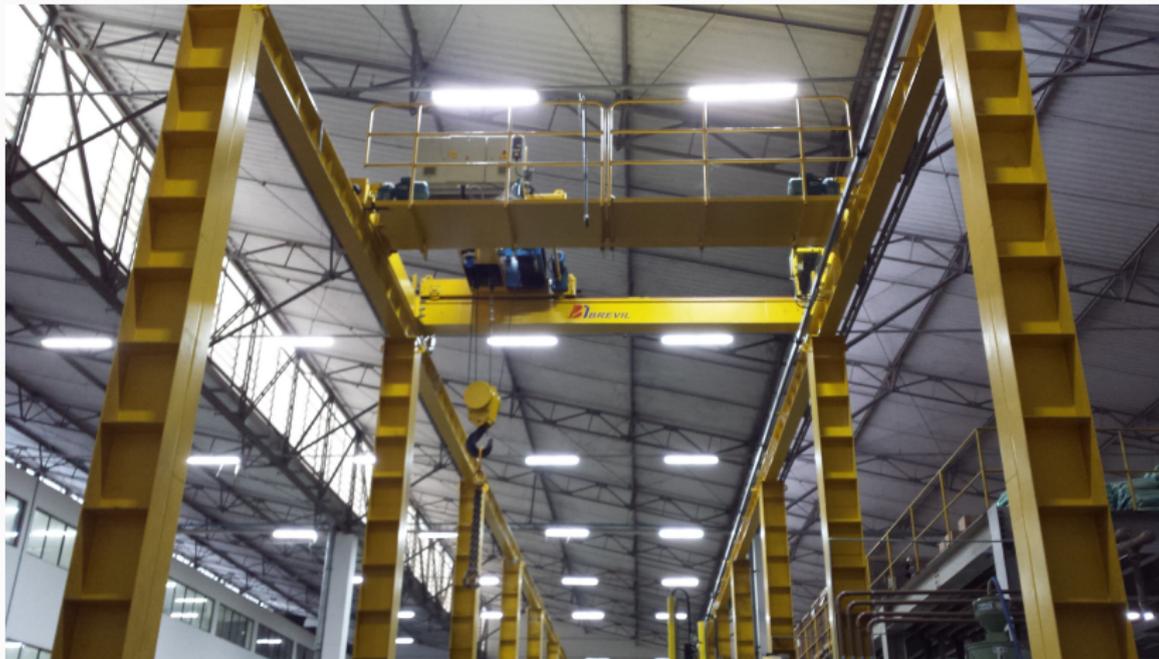
Indústria de eletrodomésticos com extrusoras de plástico.

Caracterização da aplicação



Moldes das extrusoras precisam ser substituídos através de uma ponte rolante.

Caracterização da aplicação



Estrutura da ponte rolante pela fábrica.

Objetivos

- Detectar pessoas automaticamente na região de trabalho.
- Impedir movimentação da ponte ao detectar pessoas.

Objetivos

- Detectar pessoas automaticamente na região de trabalho.
- Impedir movimentação da ponte ao detectar pessoas.

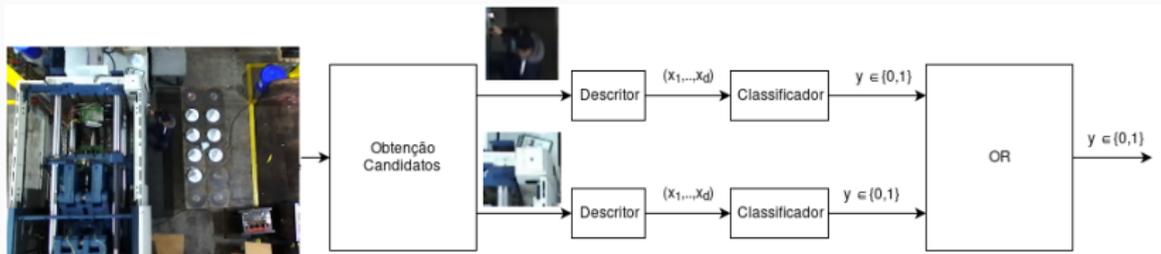
Funcionamento

- Câmera de profundidade com vista superior da área de trabalho.
- Aprendizado de máquina e visão computacional para detecção de pessoas.

Métodos de detecção de objetos

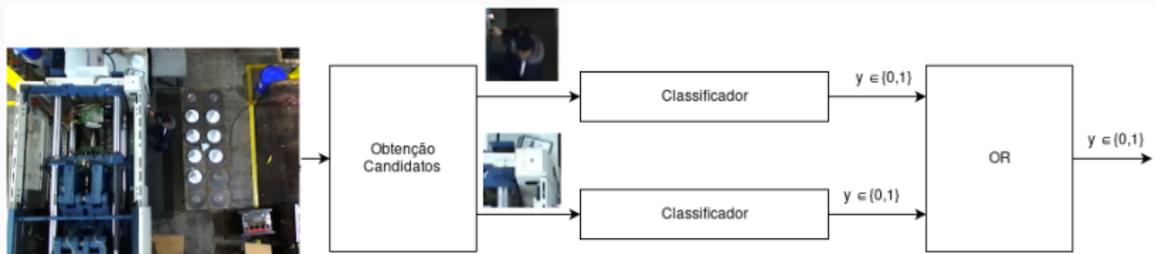
Utilizando descritores

1. Identificar candidatos.
2. Utilizar um extrator de características para descrição do objeto.
3. Introduzir a amostra, proveniente do descritor, em um classificador.



Utilizando aprendizado de representação

1. Identificar candidatos na imagem.
2. Introduzir a amostra em um classificador profundo, obtendo a classe correspondente.



- Detecção utilizando métodos tradicionais de aprendizado.
- Classificação utilizando aprendizado de representação.
- Resultados

Detecção utilizando métodos tradicionais

Michael Rauter. 2013. Reliable Human Detection and Tracking in Top-View Depth Images. In *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW '13)*.

- Detecção de candidatos baseados em imagem de profundidade.
- Descritores de características baseado em blocos simples.
- Classificador SVM.
- Ambiente controlado.

Imagens de profundidade



Imagem de cor.

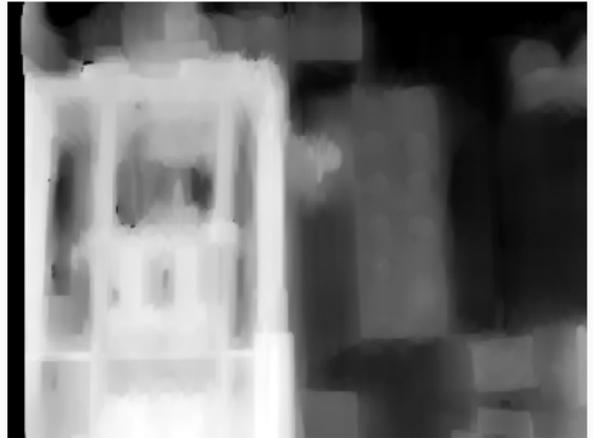


Imagem de profundidade.

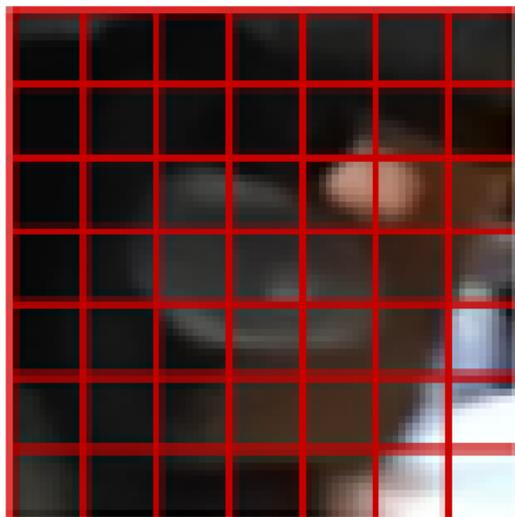
Obtenção de candidatos

- Hipótese: pessoas estão entre os objetos mais altos da cena.
- Solução: obter máximos locais.
- Centralização com *mean shift*. (CHENG, 1995).

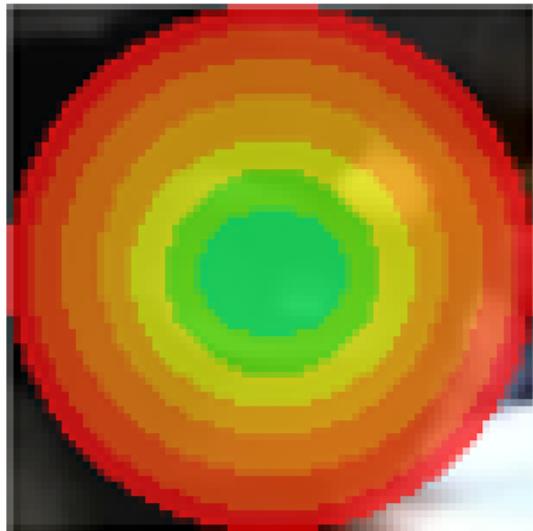


Descritores de características

Grades simples



Aneis circulares



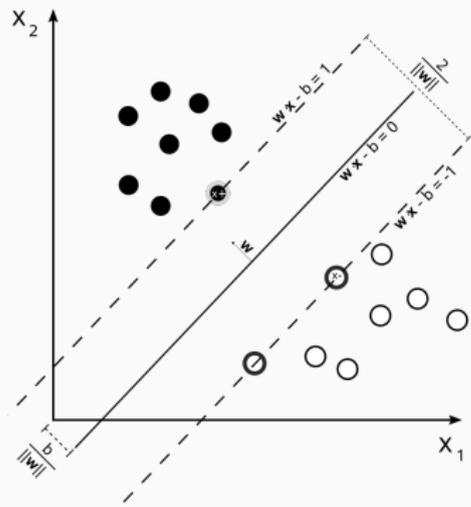
Classificador *Support Vector Machine* binário

A função decisão é dada por

$$f(x) = w^T x + b.$$

Otimização da função custo

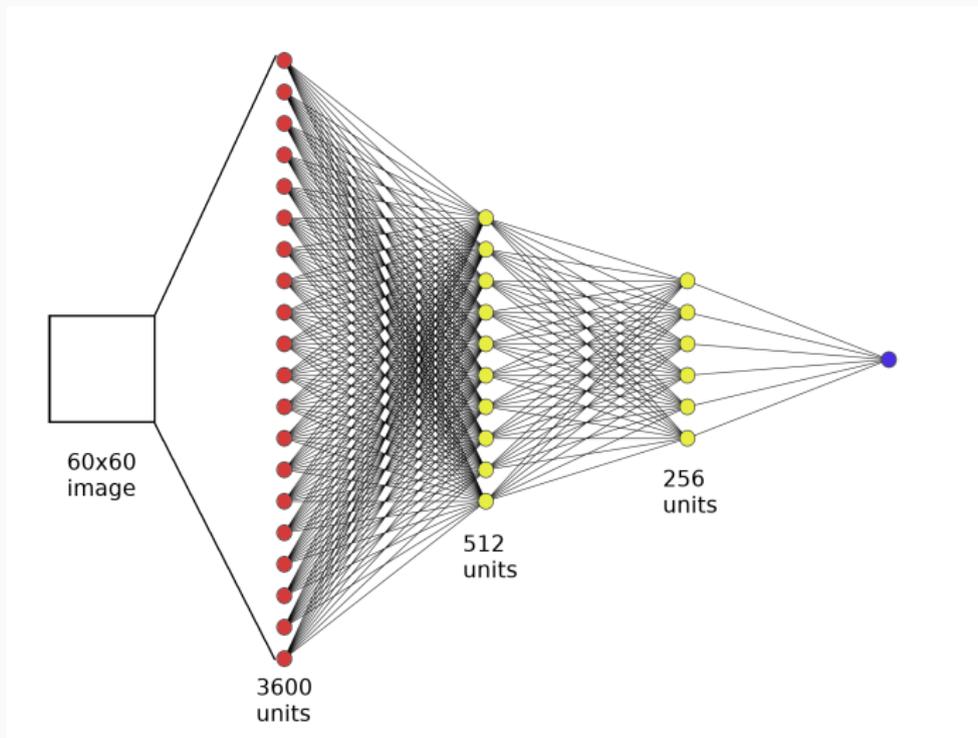
$$\min_{w,b} P(w, b) = \frac{1}{2} \|w\|^2 + C \sum_i H_1(y_i f(x_i)),$$



- Maximização da margem *versus* erros de treinamento.
- Transformação de espaço: Kernel RBF, parâmetro σ .
- Escolha hiper-parâmetros: Validação Cruzada (5 conjuntos).

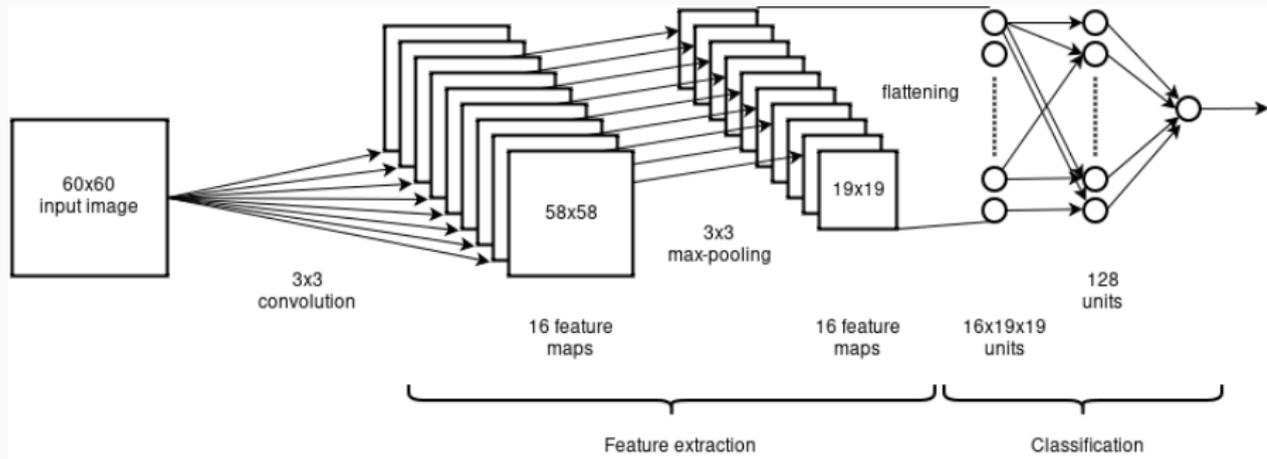
Classificação utilizando aprendizado de representação

Rede neural densamente ligada (MLP)



1.975.297 parâmetros

Rede neural convolucional (CNN)



739.745 parâmetros

- Otimização (não convexa) da função entropia cruzada binária (*binary cross entropy*)

$$C(\theta) = -\frac{1}{N} \sum_i^N [y_i \ln f(x_i, \theta) + (1 - y_i) \ln(1 - f(x_i, \theta))]$$

- Utiliza-se o algoritmo de otimização para mini-batch SGD Adam.
- Não utilizou-se regularização.

Conjunto de datos

- Vídeos gravados durante uma única oportunidade de visita à fábrica.
- Iluminação padrão durante uma manhã de atividade normal.
- Máquinas se repetem mas não pessoas.

Treinamento

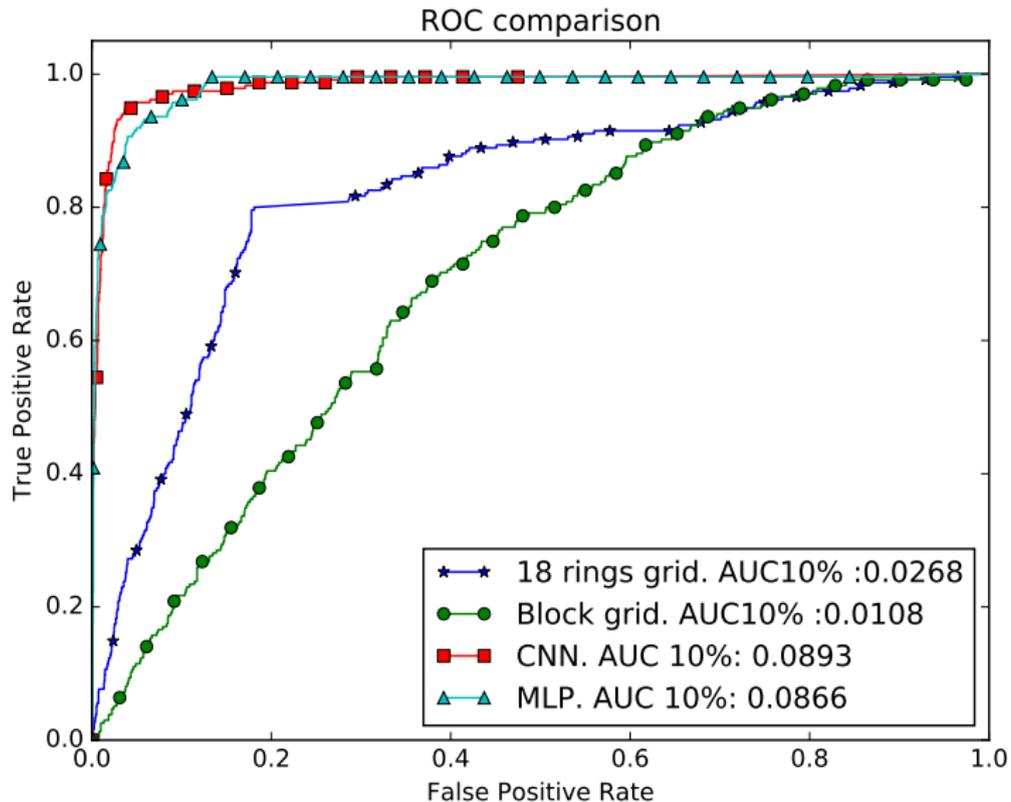
- Candidatos extraídos e manualmente classificados.
- 16898 amostras: 14966 negativas e 1932 positivas.
- Forte desbalanceamento foi um desafio para redes neurais.

Testes

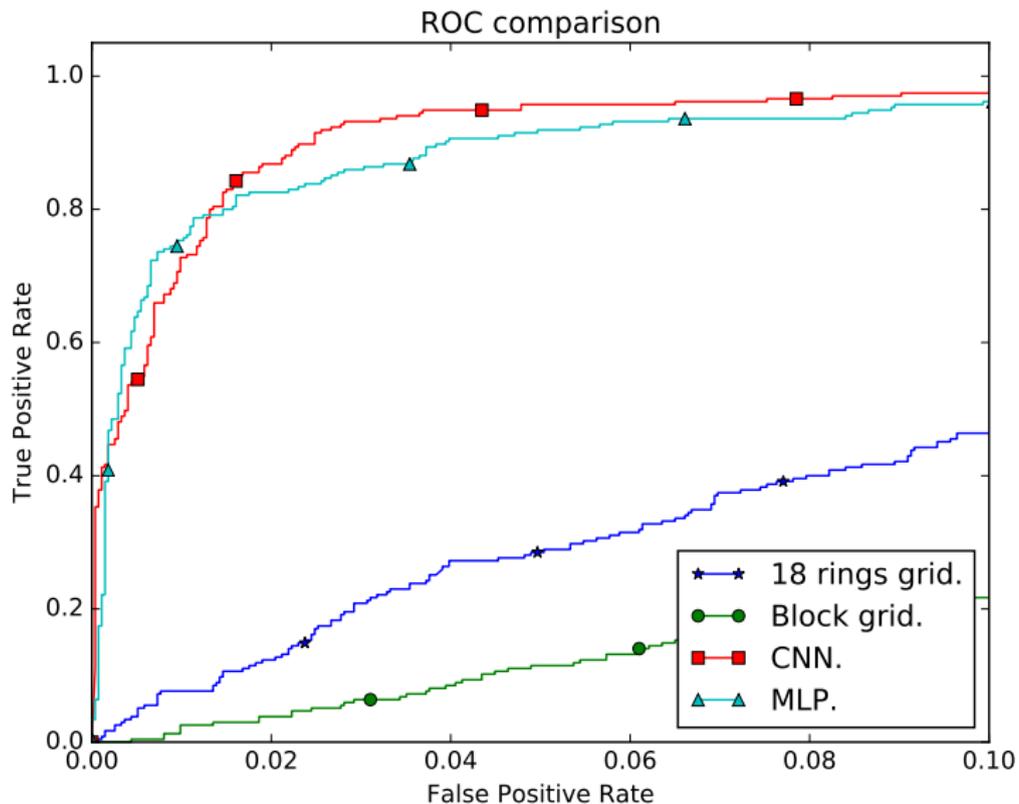
- Candidatos extraídos de dois vídeos remanescentes.
- 2973 amostras: 2738 negativas e 235 positivas.

Resultados

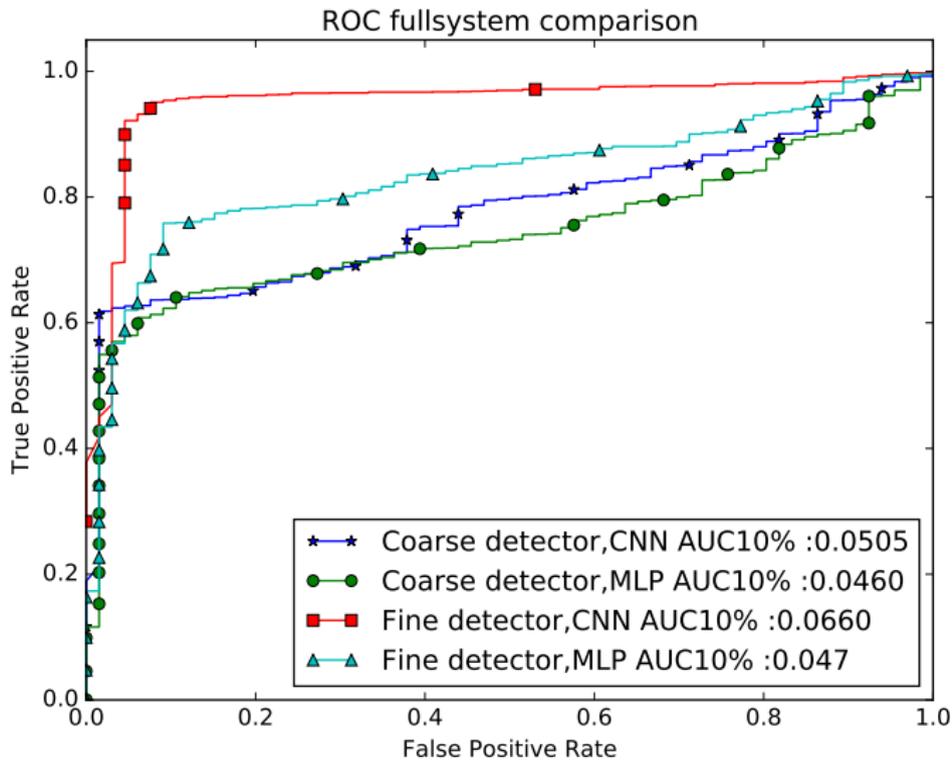
Classificadores



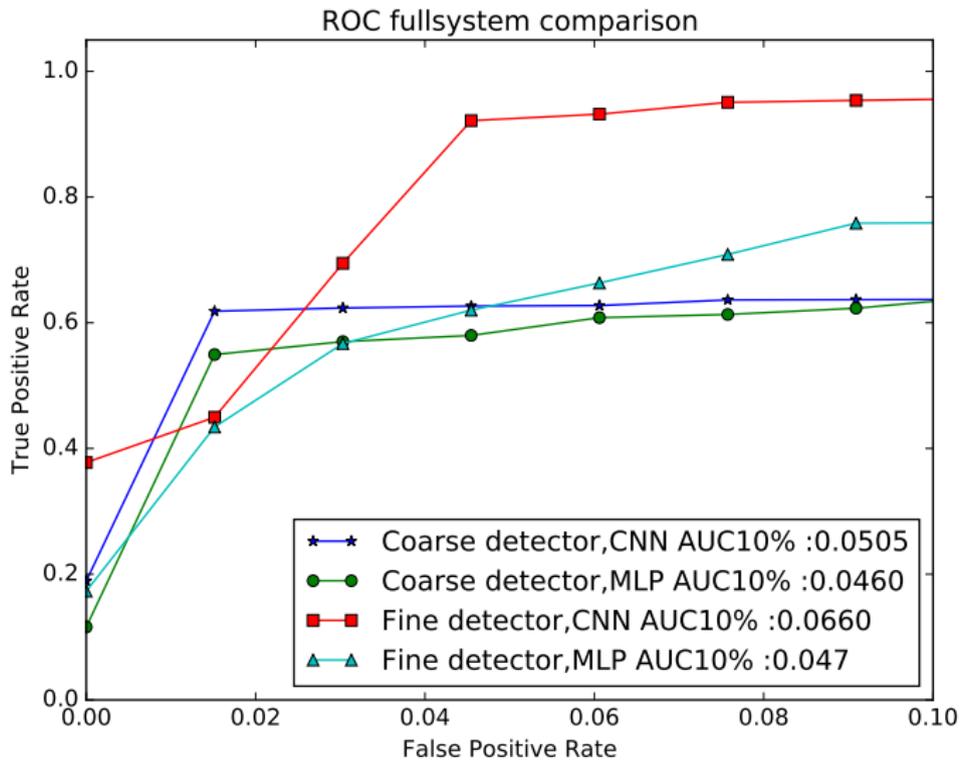
Classificadores



Sistema completo



Sistema completo



Conclusão

Conclusão

- Sistema de detecção de pessoas em um ambiente industrial.
- Duas propostas de solução
 - Métodos tradicionais de aprendizado
 - Métodos de aprendizado de representação
- Desempenho superior das técnicas de aprendizado de representação em relação aos métodos tradicionais de aprendizado.
- Bom desempenho mesmo com conjuntos medianos e desbalanceados.
- Modelos profundos demandam mais tempo de treinamento (1.5h GPU vs <1 min CPU).

Obrigado!

Contato

eduardohenriquearnold.github.io