

Técnicas de aprendizado de máquina para detecção de pessoas em ambiente industrial

Eduardo Henrique Arnold

7 de Dezembro de 2016

Universidade Federal de Santa Catarina

Introdução

Caracterização da aplicação

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.

Sistema de segurança

Objetivos

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

Sistema de segurança

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.

Introdução

Aprendizado de máquina

Métodos tradicionais de aprendizado

- Requerem pré-processamento das amostras: descrição.
- Criar um extrator de características requer conhecimento do problema.
- A descrição da amostra é crítica, em geral tem mais ênfase que o modelo em si.
- Modelos comuns: SVM, redes neurais, *random forests*.

Métodos de aprendizado profundo

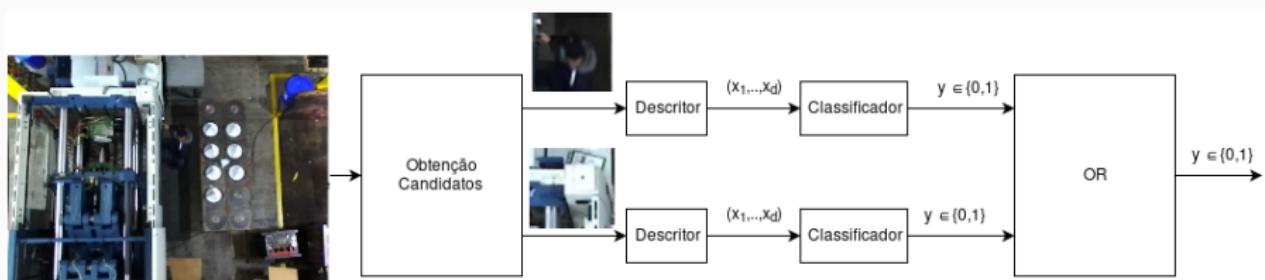
- Tem como entrada dados “crus”.
- Aprendizado de representação.
- Modelo com muitos parâmetros e diversas camadas.
- Requer um grande número de amostras para o treinamento.
- Inovações que permitem treinamento.
- Modelos comuns: Redes neurais profundas, convolucionais, recursivas (LSTM).

Introdução

Método de detecção de objetos

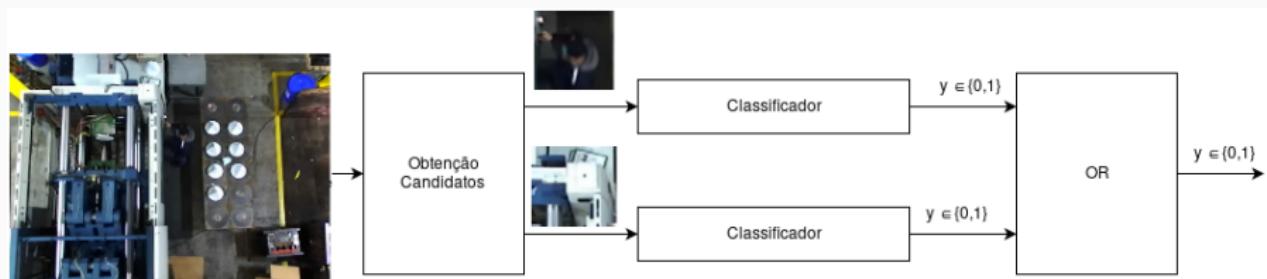
Utilizando métodos tradicionais

1. Identificar candidatos na imagem.
2. Utilizar um extrator de características para descrição do objeto.
3. Introduzir a amostra, proveniente do descriptor, em um classificador SVM, obtendo a classe correspondente.



Utilizando técnicas profundas

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



Panorama do trabalho

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

Detecção utilizando métodos tradicionais

Obtenção de candidatos

Obtenção de candidatos

- Utilizar informação da profundidade.



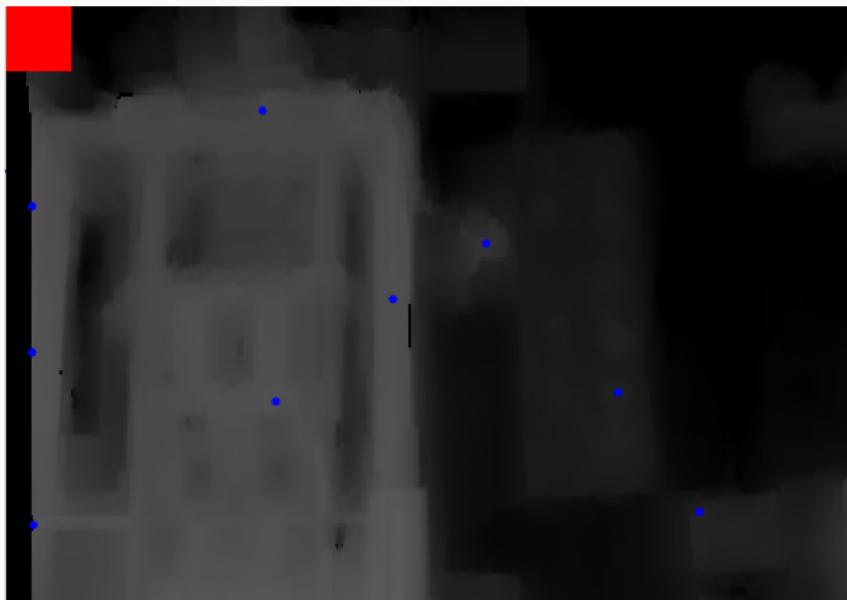
Imagen de cor.



Imagen de profundidad.

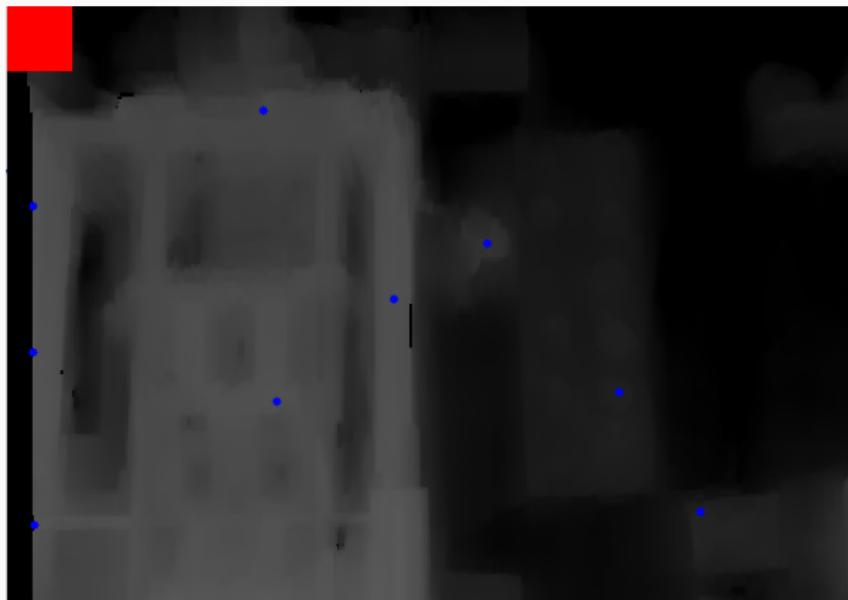
Obtenção de candidatos

- Utilizar informação da profundidade.
- Hipótese: pessoas estão entre os objetos mais altos da cena.
- Solução: obter máximos locais.



Obtenção de candidatos

- Utilizar informação da profundidade.
- Hipótese: pessoas estão entre os objetos mais altos da cena.
- Solução: obter máximos locais.
- Desafio: tamanho ideal da janela?



Obtenção de candidatos

- Estima-se tamanho do quadrado baseado na altura do ponto de máximo.
- Utiliza-se *mean shift* para centralizar o quadrado sob o candidato.
- O candidato é considerado como o recorte do quadrado sob a imagem original.

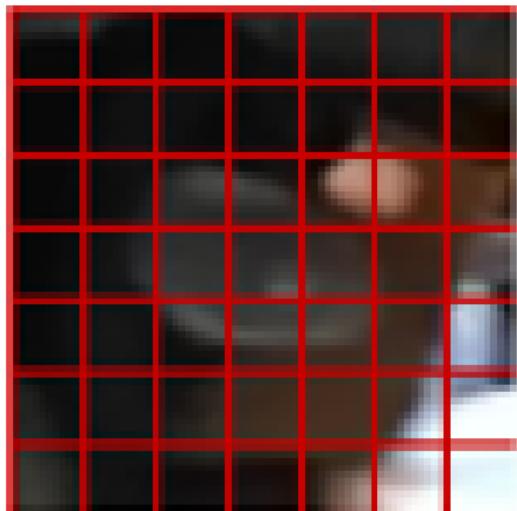


Detecção utilizando métodos tradicionais

Descritores de características

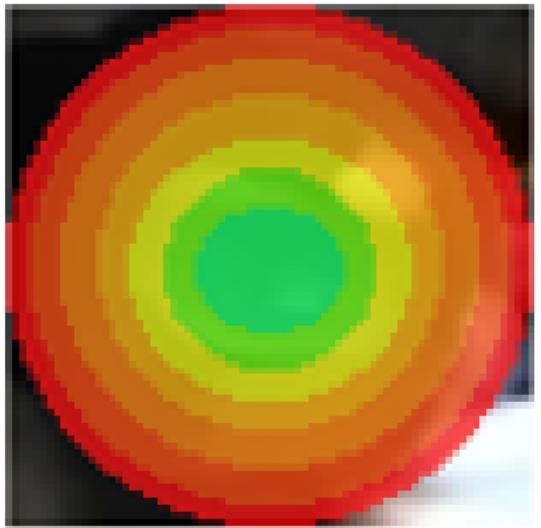
Grades simples

1. Dividir o candidato em 7×7 blocos iguais.
2. Obter média dos pixels de cada bloco \rightarrow Matriz de médias 7×7 .
3. Subtrair dessa matriz a média do bloco central.
4. Gerar histograma da matriz resultante, $x \in \mathbb{R}^d$, com $d = 32$ intervalos.



Anéis concêntricos

1. Dividir o candidato em n coroas circulares com espaçamento uniforme.
2. Obter vetor de médias dos pixels de cada coroa $\rightarrow m \in \mathbb{R}^n$.
3. Subtrair desse vetor a média da coroa interna, m_0 .
4. Diferenciar o vetor resultante, resultando em $x \in \mathbb{R}^d$ com $d = n - 1$.



Detecção utilizando métodos tradicionais

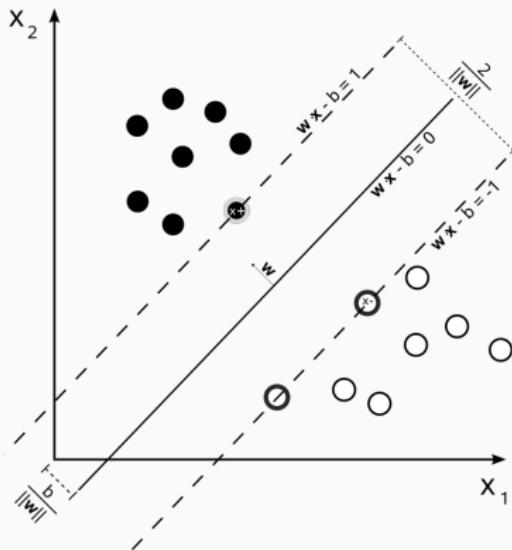
Classificação

Support Vector Machine binário

Seja o dataset o conjunto (x_i, y_i) para $i = 1 \dots N$ com $x_i \in \mathbb{R}^d$ e $y_i \in \{-1, 1\}$. Sejam os parâmetros do modelo $w \in \mathbb{R}^d$ e $b \in \mathbb{R}$.

A função decisão é dada por

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

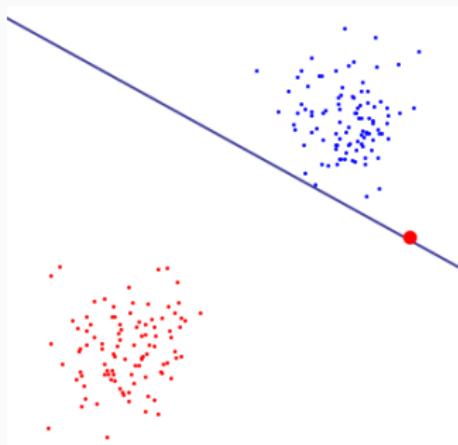


Treinamento

O treinamento consiste na otimização (convexa) 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.

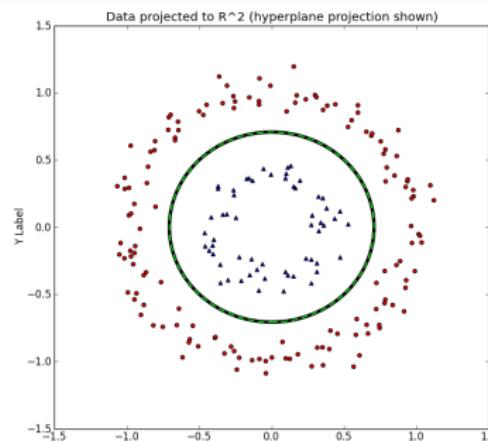
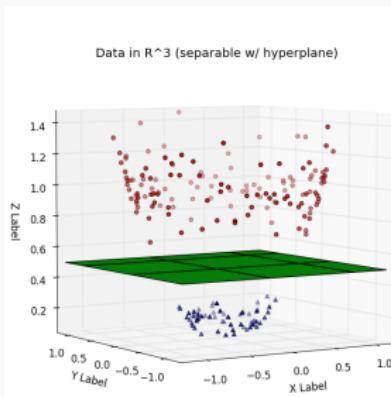


Treinamento

O treinamento consiste na otimização (convexa) 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 .



Treinamento

O treinamento consiste na otimização (convexa) da função custo

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

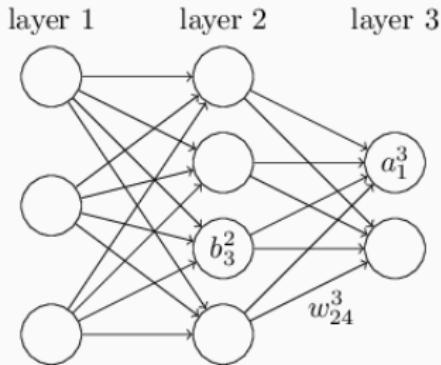
- Maximização da margem *versus* erros de treinamento.
- Transformação de espaço: Kernel RBF, parâmetro σ .
- Escolha hiper-parâmetros: Validação Cruzada (5 conjuntos).

Classificação utilizando aprendizado profundo

Redes neurais

Rede neural artificial

Também chamada de Rede perceptron multicamadas (MLP).



A ativação (saída) da unidade j na camada l é dada por

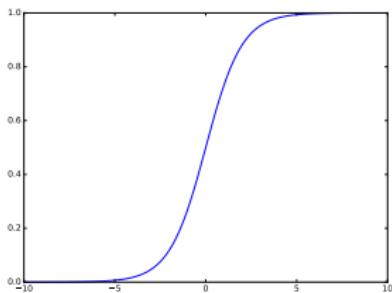
$$a_j^l = \phi \left(\sum_k w_{jk}^l a_k^{l-1} + b_j^l \right). \quad (1)$$

Funções de ativação

Sigmoide

- Historicamente utilizada para toda rede.
- Saturação nas extremidades.
- Saída no intervalo $[0, 1]$, interpretação probabilística.

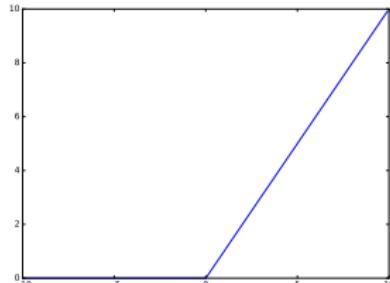
$$\sigma(z) = \frac{1}{1 + e^{-z}}.$$



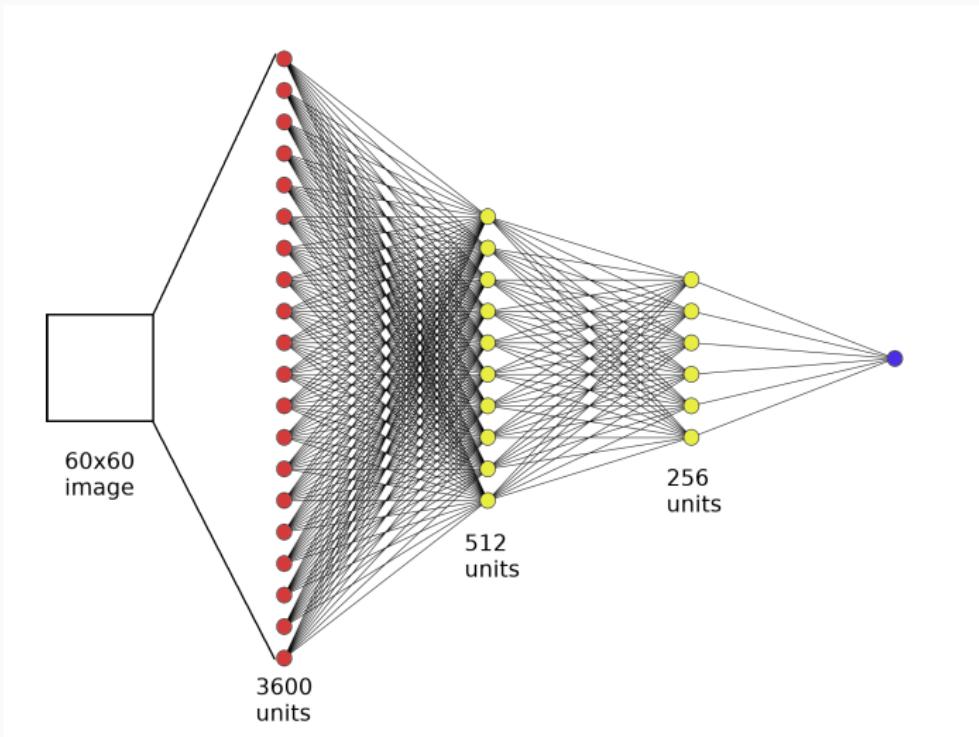
Rectified Linear Unit (ReLU)

- Derivada constante.
- Usada nas camadas intermediárias.
- Permite treinamento em redes profundas.

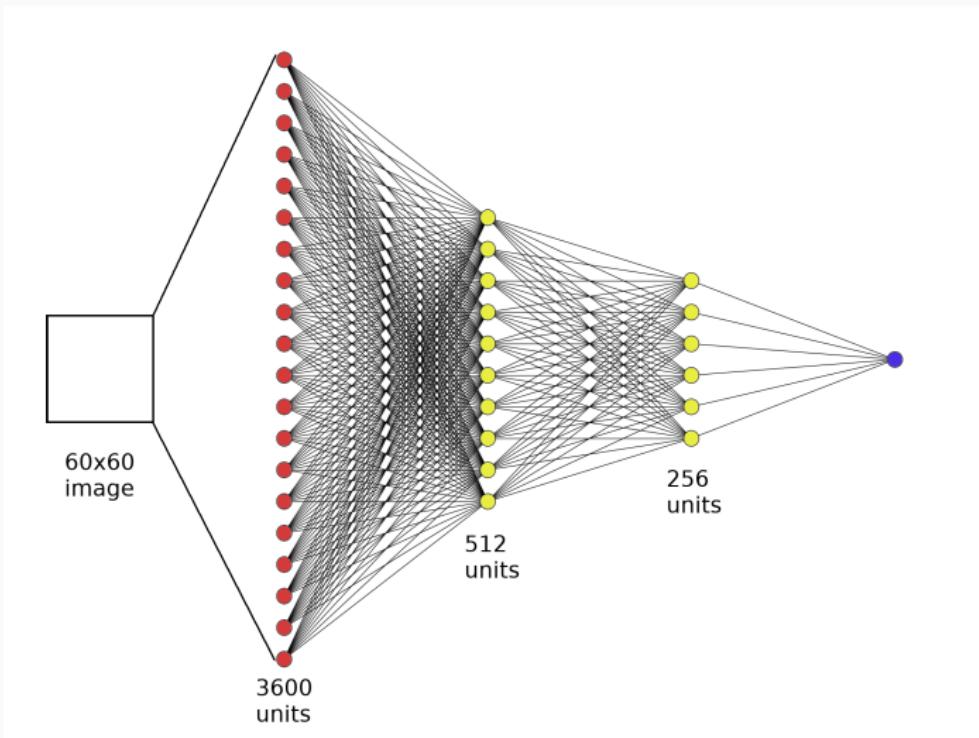
$$\text{ReLU}(z) = \max\{0, z\}.$$



Rede neural artificial (MLP)



Rede neural artificial (MLP)



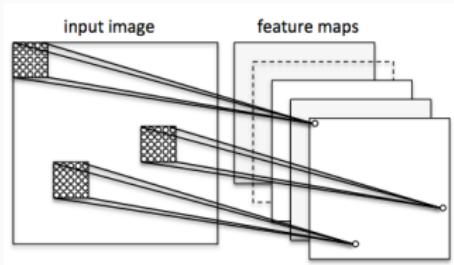
Aprendizado profundo?

Classificação utilizando aprendizado profundo

Redes convolucionais

Redes convolucionais

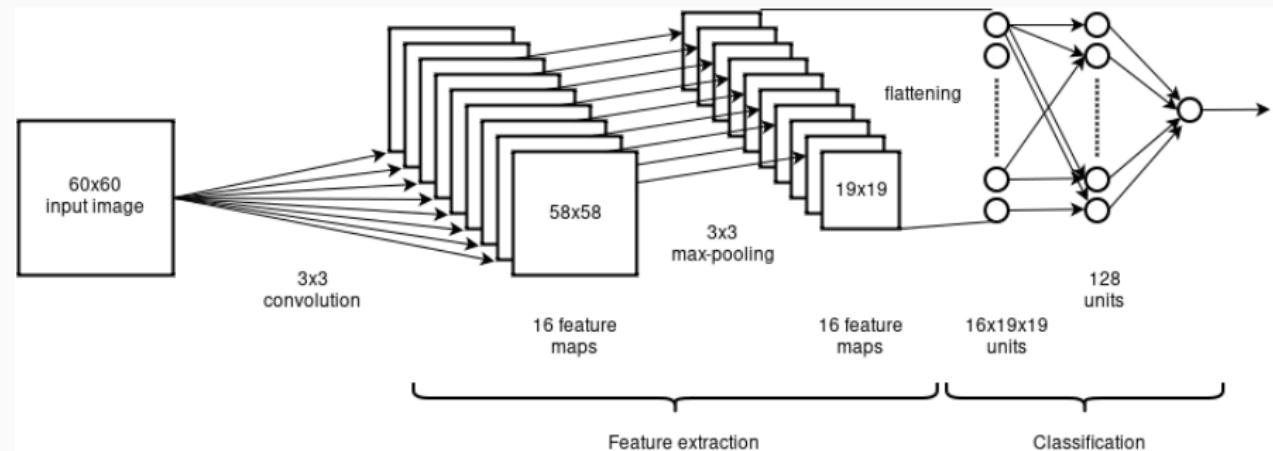
- Redes convolucionais são uma variação das redes neurais.
- Os nós são ligados através de campos receptivos (grupo de pixels adjacentes).
- Pesos compartilhados entre campos receptivos, redução dos parâmetros.
- Permite invariância à translação.



Pode-se observar essa operação como uma convolução:

$$a^1 = \phi(w * a^0 + b).$$

Redes convolucionais(CNN)



Classificação utilizando aprendizado profundo

Processo de treinamento

Função custo

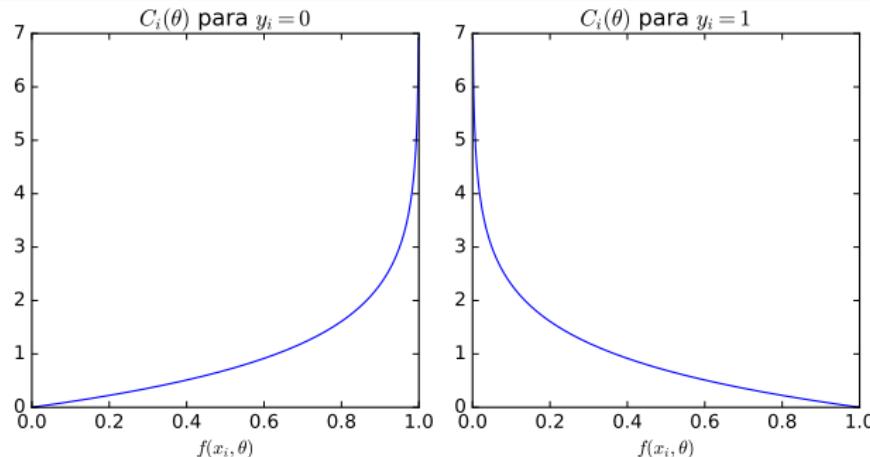
- Indica o desempenho do modelo conjunto de treinamento.
- Para um modelo ideal teria o valor nulo.

Função custo

- Indica o desempenho do modelo conjunto de treinamento.
- Para um modelo ideal teria o valor nulo.

Utiliza-se a função entropia cruzada binária (*binary cross entropy*)

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



Otimização

- Minimizar a função custo é um problema de otimização não convexa.
- Utiliza-se o algoritmo *Mini-Batch Gradient Descent*.
- MB-GD é uma variação que avalia o gradiente em um conjunto de amostras ($1 < k \leq N$) a cada iteração.
- É necessário calcular o gradiente da função custo.

Otimização

- Minimizar a função custo é um problema de otimização não convexa.
- Utiliza-se o algoritmo *Mini-Batch Gradient Descent*.
- MB-GD é uma variação que avalia o gradiente em um conjunto de amostras ($1 < k \leq N$) a cada iteração.
- É necessário calcular o gradiente da função custo.

Propagação reversa de erros - Backpropagation

- O cálculo do gradiente requer as derivadas parciais de cada parâmetro.
- Numa rede profunda o cálculo simbólico se torna inviável.
- Esse algoritmo faz uma propagação reversa do erro na saída para encontrar as derivadas parciais simultaneamente.

Resultados

Medidas de desempenho

Medidas de desempenho

- Classificadores possuem saída probabilística.
- Escolhe-se T , limiar de probabilidade da classe positiva (pessoa).

Medidas de desempenho

- Classificadores possuem saída probabilística.
- Escolhe-se T , limiar de probabilidade da classe positiva (pessoa).
- Cada escolha gera uma *matriz de confusão*

		Classificado	
		0	1
Real	0	TN	FP
	1	FN	TP

Matriz de confusão

Medidas de desempenho

- Classificadores possuem saída probabilística.
- Escolhe-se T , limiar de probabilidade da classe positiva (pessoa).
- Cada escolha gera uma *matriz de confusão*
- TVP e TFP formam ponto no espaço *Receiver Operating Characteristic*.

		Classificado	
		0	1
Real	0	TN	FP
	1	FN	TP

Matriz de confusão

$$\text{TVP} = \frac{\text{TP}}{\text{FN} + \text{TP}}$$

$$\text{TFP} = \frac{\text{FP}}{\text{TN} + \text{FP}}$$

Medidas de desempenho

- Classificadores possuem saída probabilística.
- Escolhe-se T , limiar de probabilidade da classe positiva (pessoa).
- Cada escolha gera uma *matriz de confusão*
- TVP e TFP formam ponto no espaço *Receiver Operating Characteristic*.
- Curva ROC: compromisso TVPxTFP (função de T).

		Classificado	
		0	1
Real	0	TN	FP
	1	FN	TP

Matriz de confusão

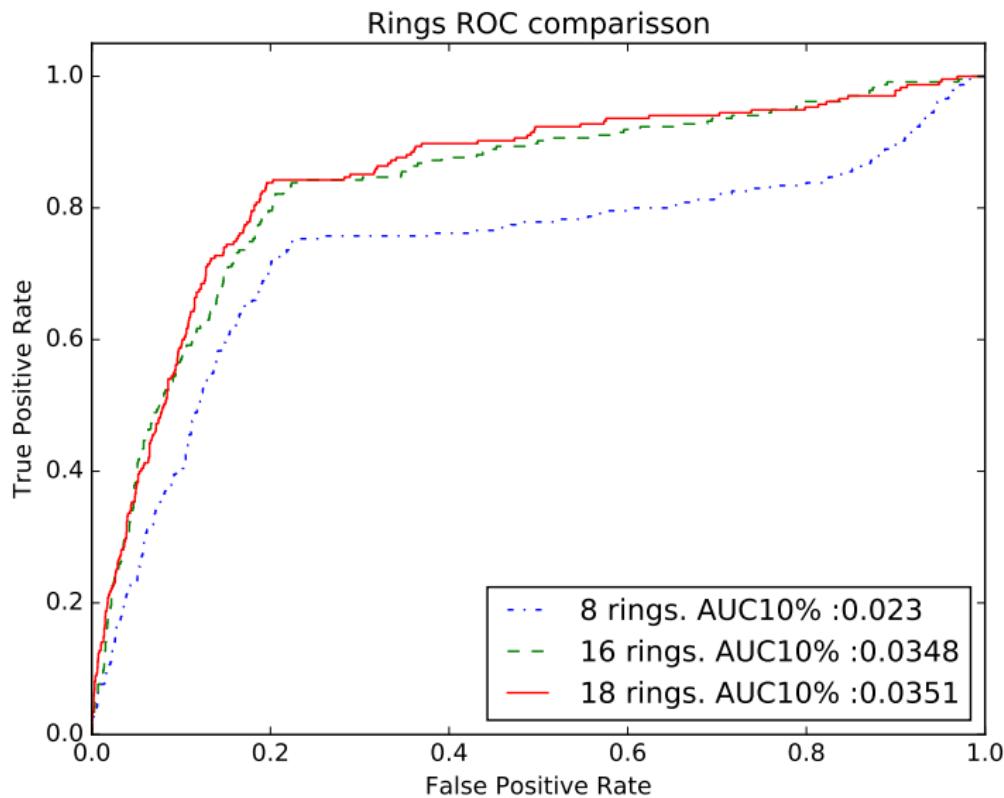
$$\text{TVP} = \frac{\text{TP}}{\text{FN} + \text{TP}}$$

$$\text{TFP} = \frac{\text{FP}}{\text{TN} + \text{FP}}$$

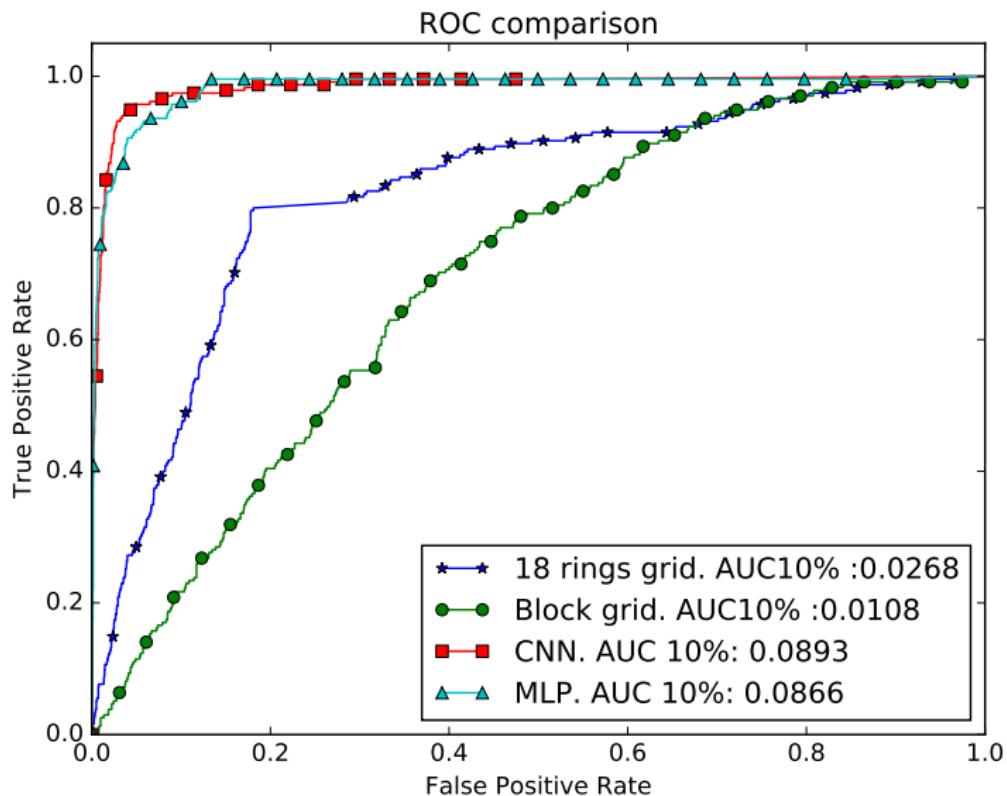
Resultados

Classificadores

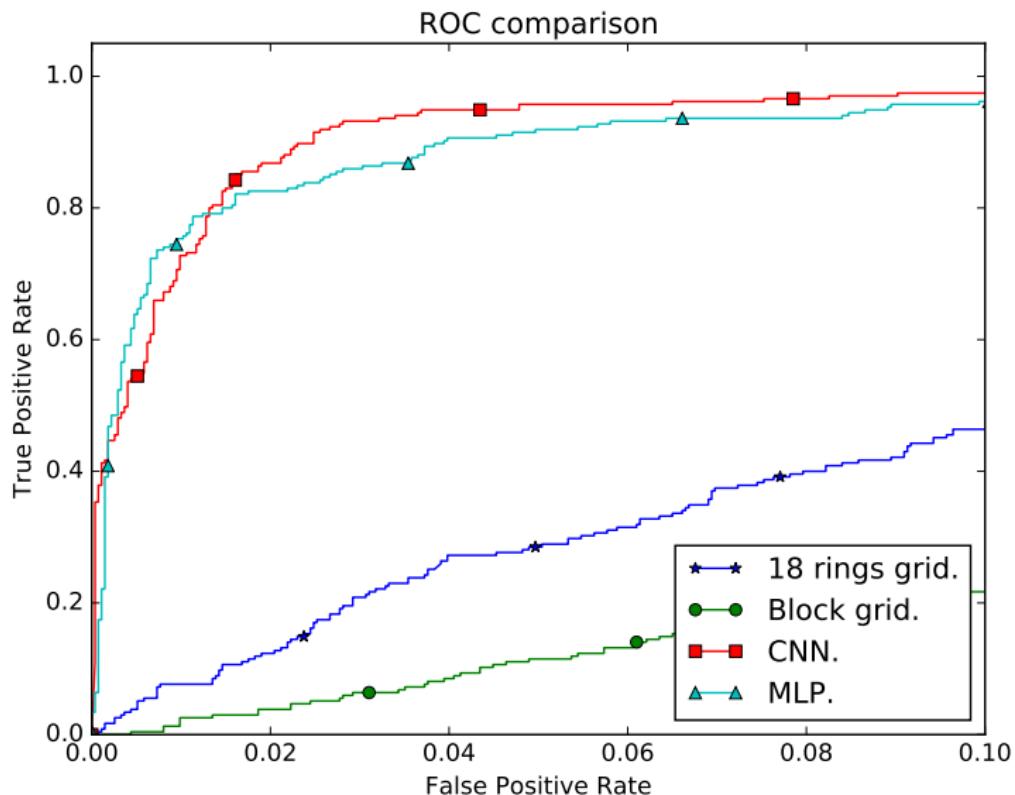
Avaliação do descriptor de anéis concêntricos



Todos os classificadores



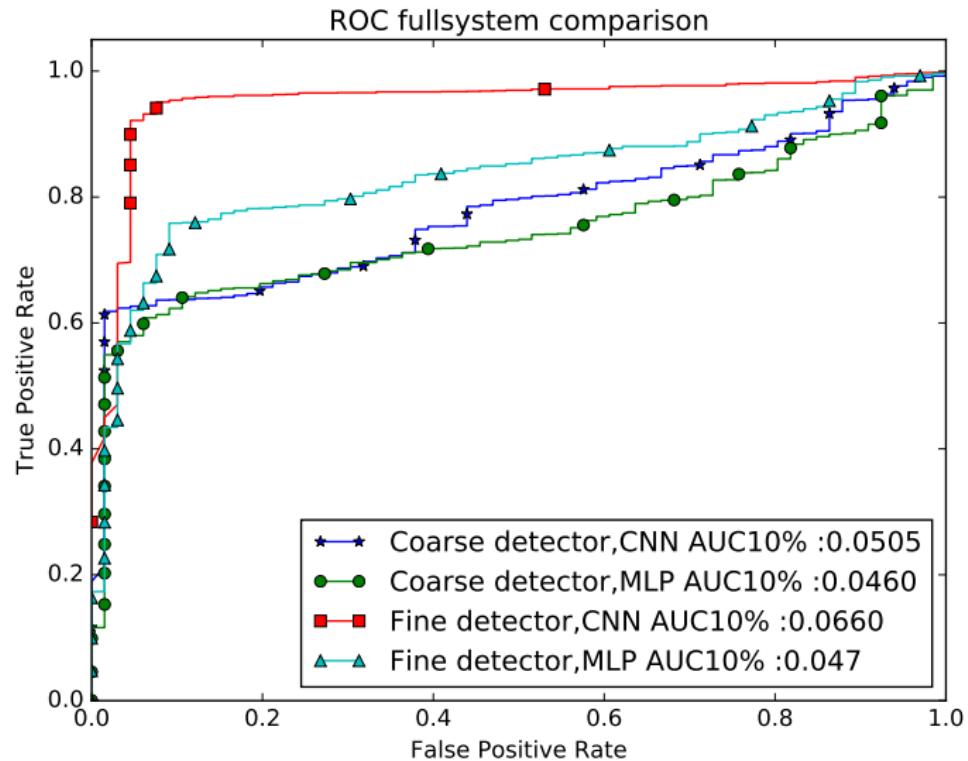
Todos os classificadores



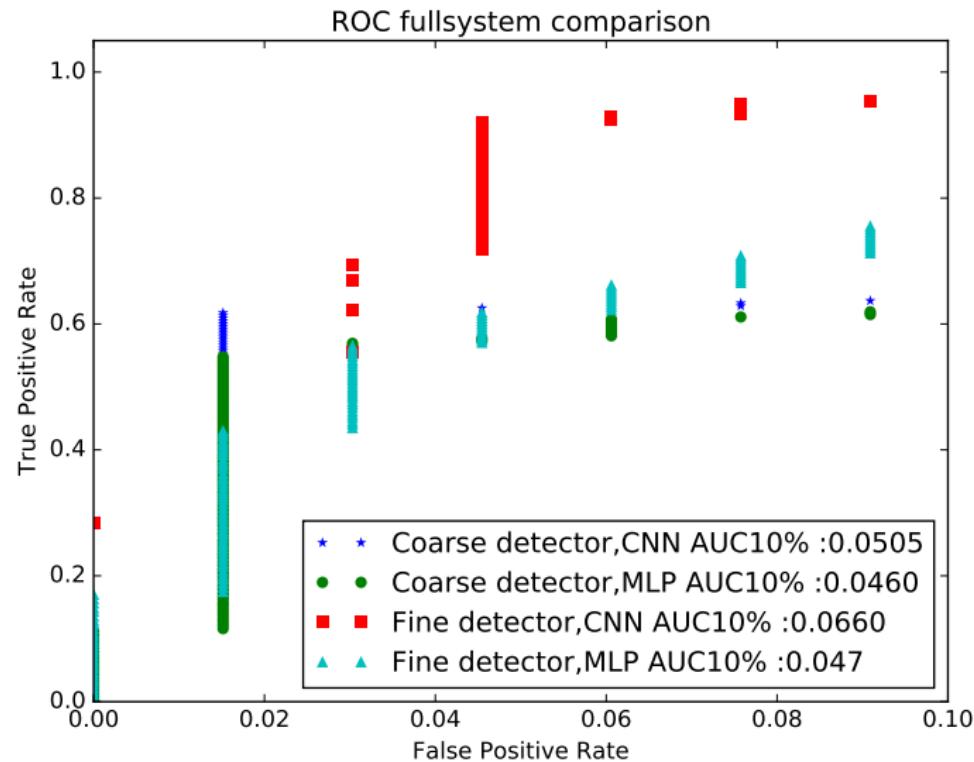
Resultados

Sistema completo

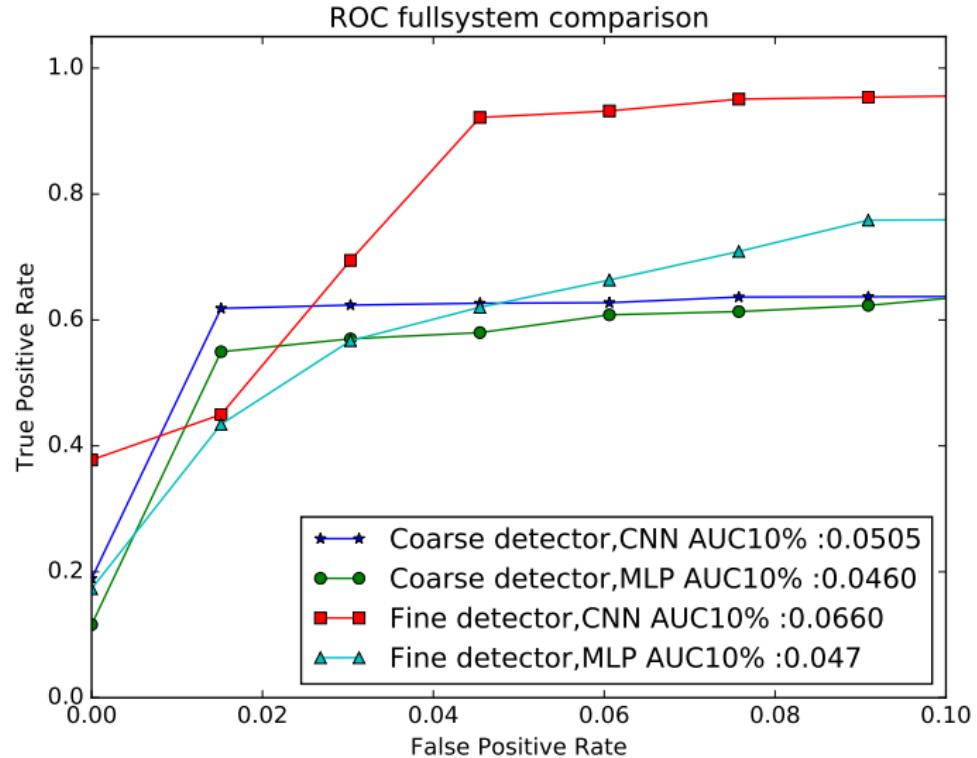
Sistema completo



Sistema completo



Sistema completo



Vídeo de demonstração

Demonstração do sistema em funcionamento no vídeo de testes.

Conclusão

Conclusão

- Sistema de detecção de pessoas em um ambiente industrial.
- Duas propostas de solução
 - Métodos tradicionais de aprendizado
 - Métodos de aprendizado profundo
- Desempenho superior das técnicas profundas em relação aos métodos tradicionais de aprendizado.
- Bom desempenho mesmo com conjuntos medianos e desbalanceados.

Trabalhos futuros

Ajuste fino do classificador

- Vídeo de testes limitado: poucos quadros sem pessoas, mesmos moldes.
- Avaliar a ocorrência de falsos positivos após instalação.
- Fazer treinamento de ajuste fino na camada de classificação.

Trabalhos futuros

Ajuste fino do classificador

- Vídeo de testes limitado: poucos quadros sem pessoas, mesmos moldes.
- Avaliar a ocorrência de falsos positivos após instalação.
- Fazer treinamento de ajuste fino na camada de classificação.

Proposta de solução integralmente profunda

- Utilizar uma rede neural convolucional.
- O quadro inteiro é tido como entrada do sistema.
- Saída bidimensional como mapa de probabilidades.
- Evita necessidade de seleção de candidatos.

Obrigado