Detecção de Objetos com Deep Learning

Trabalho final da disciplina PSI5886 - Princípios de Neurocomputação

Diego Cardoso - *diegocardoso@usp.br*
Vitor Finotti - *vfinotti@usp.br*
Vitor Hugo Meneses Beck - *vmenesesbeck@usp.br*

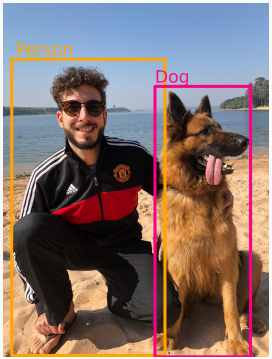
Universidade de São Paulo
Escola Politécnica

Agenda

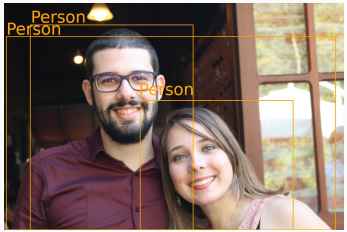
1. O grupo
2. Introdução
3. Detecção de objetos com deep learning
4. Ensaios experimentais
5. Conclusões

O grupo

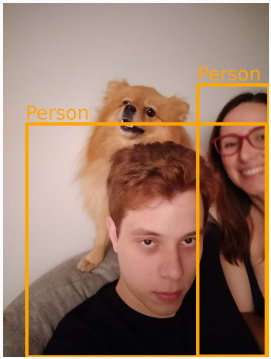
Membros



(a) Diego Cardoso



(b) Vitor Hugo Meneses Beck



(c) Vitor Finotti

Figura 1: Membros do grupo, classificados pelo algoritmo treinado. Autoria própria.

Divisão de tarefas

	Diego C.	Vitor F.	Vitor H.
Intro - Detecção Objetos	X	X	
Intro - Algoritmos Clássicos	X		
Intro - CNN		X	
R-CNN	X		
Fast R-CNN	X		
Faster R-CNN		X	X
YOLO		X	X
Ensaios experimentais			X
Conclusões	X	X	X

Tabela 1: Divisão de tarefas e responsabilidades do grupo.

Introdução

A detecção de objetos

- Dois estágios
  - Localização de objetos
  - Classificação dos objetos
- PASCAL VOC Challenge [1]: Competição de deteção de objetos
- Abordagens
  - Clássicas
  - *Deep learning*

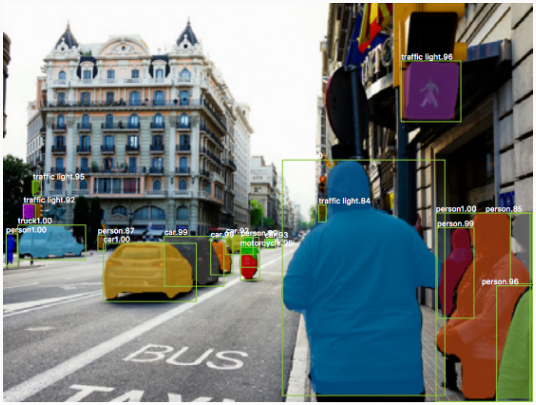


Figura 2: Exemplo de detecção de objetos. Obtido de He et al. [2].

Mas como ganhar dinheiro com isso?

- Detecção de face
- Contagem de pessoas/público
- Veículos Autônomos
- Contagem industrial
- Segurança
- Mapeamento de terreno
- Controle de qualidade

Figura 3: Exemplo de aplicação de detecção de objetos pra controle de qualidade.

Linha do Tempo

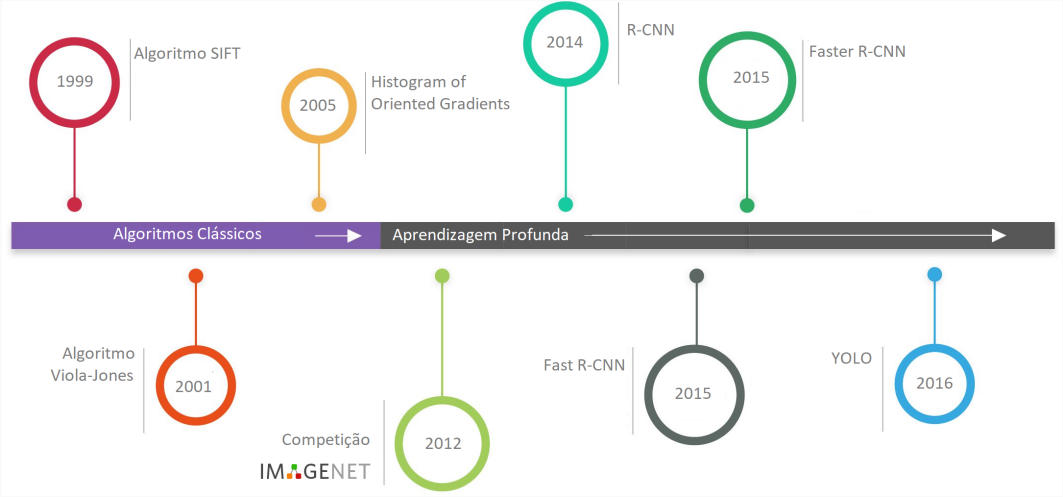


Figura 4: Linha do tempo da evolução dos algoritmos de detecção de objetos.

Algoritmos clássicos

- Algoritmo Viola-Jones [3]
  - Cascata de classificadores "fracos"
- Algoritmo SIFT (*Scale Invariante Feature Transform*) [4]
  - Correspondência entre ponto e imagem
  - Invariante à rotação, translação e dimensionamento
- HOG (*Histogram of Oriented Gradients*) [5]
  - Robusto à mudanças de iluminação e cor
  - Objetos descritos pela distribuição do gradiente de intensidade dos pixels e direções das bordas

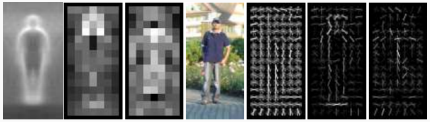


Figura 5: Exemplo de aplicação do descritor HOG [5].

Redes neurais convolucionais na visão computacional

- Proposto por LeCun et al. [6]
- Destaque após ImageNet 2012 [7]
  - Redução do erro de classificação pela **metade**!
- Arquitetura
  - **Camada Convolucional:** seleção de características
  - **Pooling:** Sub-amostragem
  - **Fully-Connected:** Combinação de características

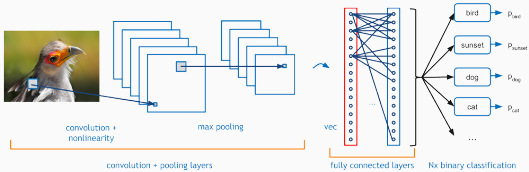


Figura 6: Exemplo de rede neural convolucional. Obtido de Deshpande [8].

Detecção de objetos com deep learning

R-CNN

- Primeira a usar *deep learning*
- Etapas
  1. *Selective Search*: proposta de regiões
  2. CNN: extração
  3. SVM: classificação
  4. Regressão para ajuste das caixas
- Desempenho
  - Iterações:
    2000
  - Inferência:
    50s/img
  - mAP(*mean Average Precision*, VOC 2012 [9]):
    49.6

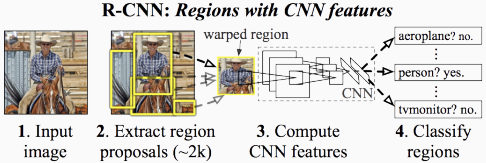


Figura 7: Rede neural convolucional baseada em regiões (R-CNN). Obtido de Girshick et al. [10].

Fast R-CNN

- Otimização do uso da CNN na *selective search*
- Extração, classificação e regressão feitos por uma única rede
- Desempenho
  - Iterações:
    2000 $\Rightarrow$ **1**
  - Inferência:
    50s/img $\Rightarrow$ **2s/img**
  - mAP (VOC 2012) [9]:
    49.6 $\Rightarrow$ **68.4**

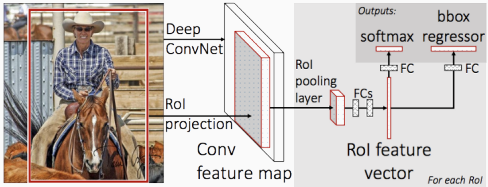


Figura 8: Arquitetura Fast R-CNN. As etapas de extração, classificação e regressão são incorporadas em uma única rede. Obtido de Girshick [11].

Faster R-CNN

- Substituição da *selective search* por uma rede neural
- Desempenho
  - Iterações:
    $2000 \Rightarrow 1 \Rightarrow 1$
  - Inferência:
    50s/img $\Rightarrow$ 2s/img $\Rightarrow$ **0.2s/img**
  - mAP (VOC 2012) [9]:
    49.6 $\Rightarrow$ 68.4 $\Rightarrow$ **70.4**

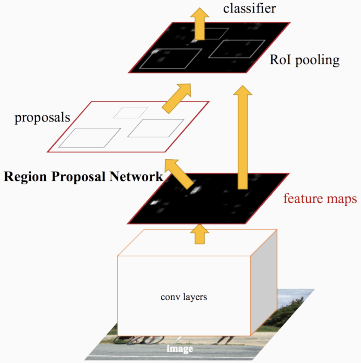


Figura 9: Faster R-CNN. Obtido de Ren et al. [12].

YOLO: You Only Look Once

- Não é baseada em regiões
- Imagem inteira é analisada
- Classificação mais rápida
- Desempenho ruim para objetos pequenos
- Desempenho
  - Iterações: 1
  - Inferência: **0.02s/img**
  - mAP (VOC 2012) [9]: 57.9

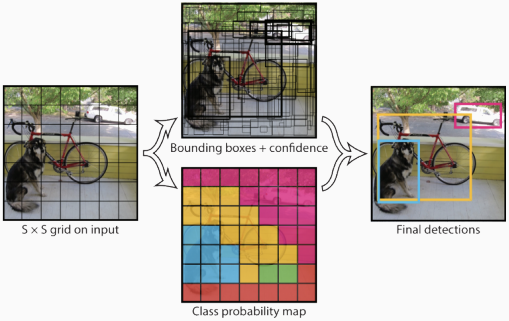


Figura 10: Arquitetura YOLO (*You Only Look Once*). Obtido de Redmon et al. [9].

Ensaios experimentais

Ensaios experimentais

- Faster R-CNN utilizada
- Resnet-50 pre treinada com ImageNet utilizada como *backbone*
- Cerca de 2h de treinamento - 20000 interações
- Hardware - Intel Core I7 8700, 8GB de RAM, NVidia 1070Ti 8GB.
- Ajuste de taxa de aprendizado após *overfitting* ocorrer.
- Código disponível em: https://github.com/roytseng-tw/Detectron.pytorch

Figura 11: Exemplo de detecção ruim. Fonte: Autoria Propria

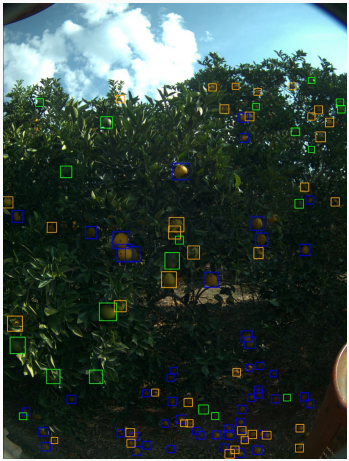
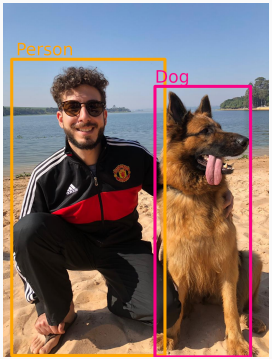
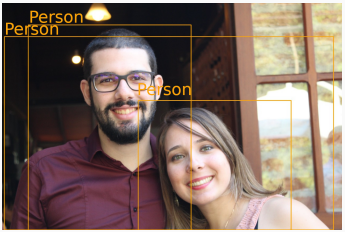


Figura 12: Exemplo de detecção boa. Fonte: Imagem cedida por Adroit Robotics.

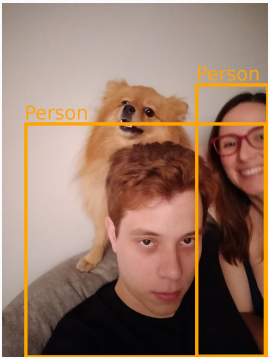
Ensaios experimentais



(a) Diego Cardoso



(b) Vitor Hugo Meneses Beck



(c) Vitor Finotti

Figura 13: Membros do grupo, classificados pelo algoritmo treinado. Autoria própria.

Conclusões

Conclusões

- *Deep learning* é benéfico
- Robustez: Faster R-CNN
- Tempo de inferência: YOLO
- Tendência da área: combinações dos métodos [9]

	R-CNN	Fast R-CNN	Faster R-CNN	YOLO
Iterações	∼2000	**1**	**1**	**1**
Inferência (s/img)	50	2	0.2	**0.02**
mAP	49.6	68.4	**70.4**	57.9

Tabela 2: Resumo do desempenho dos algoritmos. Melhores resultados em negrito. Dados obtidos de Redmon et al. [9].

Perguntas?

Referências

[1] Mark Everingham et al. “The pascal visual object classes (voc) challenge”. Em: *International journal of computer vision* 88.2 (2010), pp. 303–338.

[2] Kaiming He et al. “Mask R-CNN”. Em: *CoRR* abs/1703.06870 (2017). arXiv: 1703.06870. URL: http://arxiv.org/abs/1703.06870.

[3] P. Viola e M. Jones. “Rapid object detection using a boosted cascade of simple features”. Em: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition.* Vol. 1. Dez. de 2001, pp. I–I. DOI: 10.1109/CVPR.2001.990517.

Referências ii

[4] D. G. Lowe. "Object recognition from local scale-invariant features". Em: *Proceedings of the Seventh IEEE International Conference on Computer Vision*. Vol. 2. Set. de 1999, 1150–1157 vol.2. DOI: 10.1109/ICCV.1999.790410.

[5] N. Dalal e B. Triggs. "Histograms of oriented gradients for human detection". Em: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Vol. 1. Jun. de 2005, 886–893 vol. 1. DOI: 10.1109/CVPR.2005.177.

[6] Yann LeCun et al. "Gradient-Based Learning Applied to Document Recognition". Em: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2323. ISSN: 00189219. DOI: 10.1109/5.726791. pmid: 15823584.

Referências iii

[7] Alex Krizhevsky, Ilya Sutskever e Geoffrey E Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". Em: *Advances in Neural Information Processing Systems 25*. Ed. por F Pereira et al. Curran Associates, Inc., 2012, pp. 1097–1105. URL: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

[8] Adit Deshpande. *A Beginner's Guide To Understanding Convolutional Neural Networks*. URL: https://adeshpande3.github.io/A-Beginner%27s-Guide-To-Understanding-Convolutional-Neural-Networks/ (acesso em 12/02/2018).

Referências iv

[9] Joseph Redmon et al. "You Only Look Once: Unified, Real-Time Object Detection". Em: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 779–788.

[10] Ross Girshick et al. "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation". Em: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014, pp. 580–587.

[11] Ross Girshick. "Fast R-Cnn". Em: *Proceedings of the IEEE International Conference on Computer Vision*. 2015, pp. 1440–1448.

[12] Shaoqing Ren et al. "Faster R-Cnn: Towards Real-Time Object Detection with Region Proposal Networks". Em: *Advances in Neural Information Processing Systems*. 2015, pp. 91–99.