



# Deep Learning for Accessibility

Detection and Segmentation of Regions of Interest for  
Sign Language Recognition Systems

---

Rúbia Guerra

Orientador: Prof. Dr. Frederico Gadelha

Coorientadora: Ma. Tamires Rezende

26 de Novembro de 2019

Escola de Engenharia, UFMG

1. Introdução
2. Fundamentos
3. Bases de Dados
4. Experimentos e Resultados
5. Considerações Finais

# Introdução

---

**466 milhões** de pessoas ao redor do mundo apresentam perda auditiva debilitante [WHO, 2019]

Kuenburg et al. [2016]: **comunicação** é uma das maiores barreiras para acesso a serviços de saúde pela comunidade surda

## NOME DA GÍRIA: 007

Significado: Pessoa esperta, inteligente, sedutora, malandra ou com boa lábia

MUNDO  
estranho



**1** Mão direita em "0", com palma voltada para a esquerda



**2** Deslizar a mão para a direita, mantendo a mesma posição em "0"



**3** Com a mão na altura do ombro, fazer o sinal de "7"

Fonte: *Hirata [2019]*

Tecnologias vêm sendo estudadas para auxiliar na comunicação entre pessoas surdas e não-surdas:

## Português-Libras

HandTalk [2019], VLibras [2019].

## Libras-Português

Problema em aberto: Almeida et al. [2014], Rezende [2016], Filho et al. [2017], Amaral et al. [2019]

Er-Rady et al. [2017]:

- Complexidade visual das línguas de sinais

Er-Rady et al. [2017]:

- Complexidade visual das línguas de sinais
- Poucas bases de dados documentadas



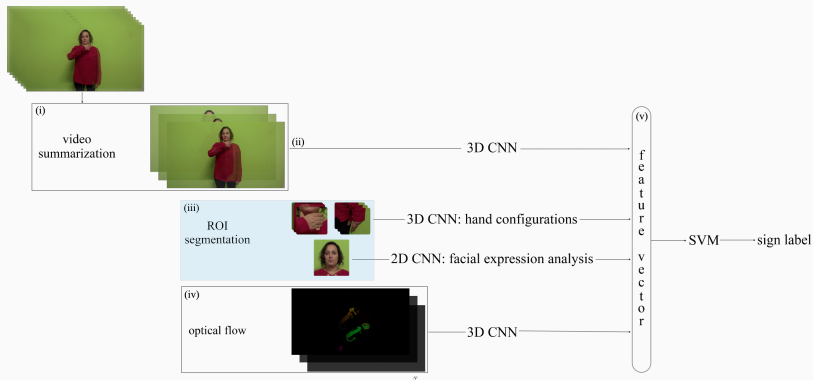
Er-Rady et al. [2017]:

- Complexidade visual das línguas de sinais
- Poucas bases de dados documentadas
- Teoria linguística pouco desenvolvida

Er-Rady et al. [2017]:

- Complexidade visual das línguas de sinais
- Poucas bases de dados documentadas
- Teoria linguística pouco desenvolvida

# Sistema de Reconhecimento de Sinais (SRS)

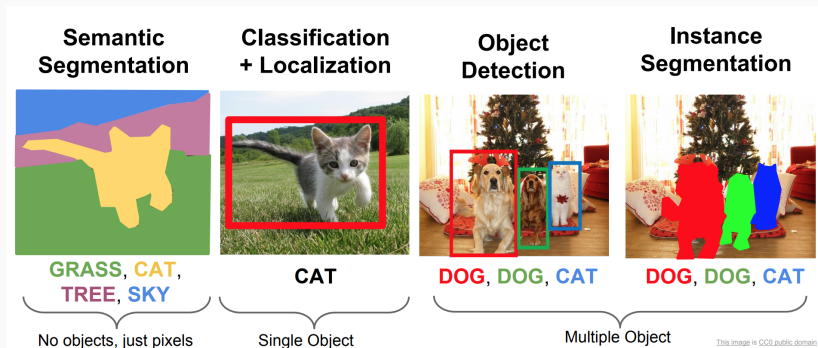


Fonte: *Beam [2017]*

# Fundamentos

---

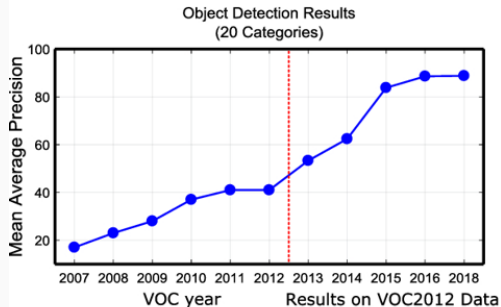
# Segmentação de Instâncias



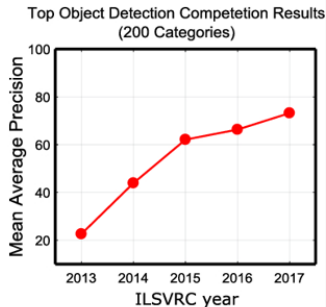
Fonte: Karpathy [2019]

# Avanços em Classificação de Imagens

Turning Point in 2012: Deep Learning Achieved Record Breaking Image Classification Result



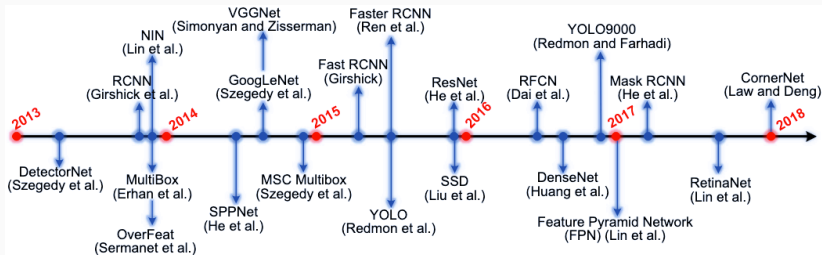
(a)



(b)

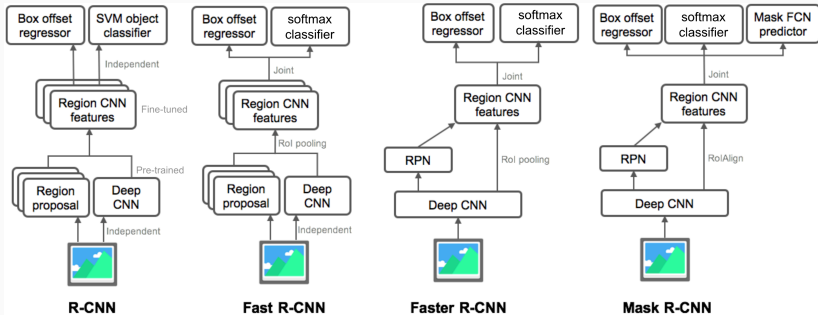
Fonte: Liu et al. [2019]

# Arquiteturas



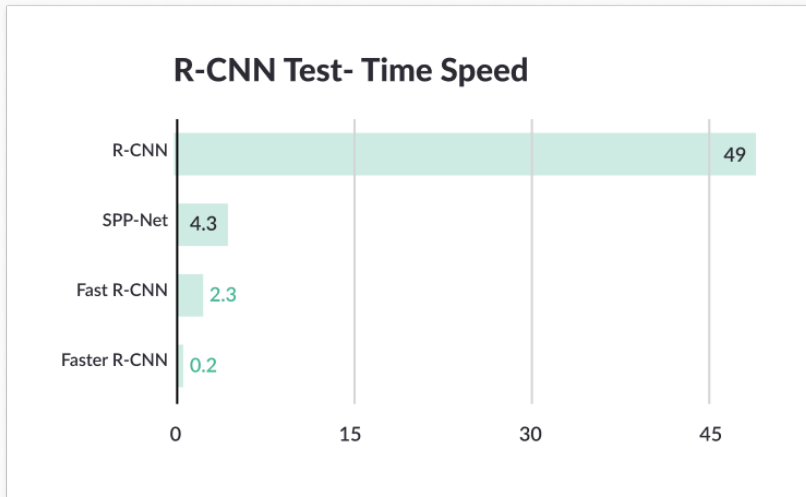
Fonte: *Liu et al. [2019]*

# Família R-CNN



Fonte: Weng [2017]





Fonte: *Gandhi [2018]*

# Bases de Datos

---

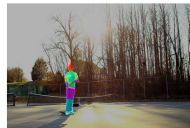
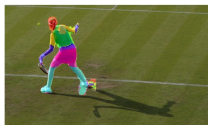
# PASCAL-Part



PASCAL-Part:  $\approx$  2 mil imagens

Fonte: *Chen et al. [2014]*

# Weakly and Semi Supervised Human Body Part Parsing via Pose-Guided Knowledge Transfer (WSHP)



WSHP:  $\approx$  150 mil imagens

Fonte: Fang et al. [2018]

Outros datasets explorados:

- EgoHands [Bambach et al., 2015]
- Rendered Handpose [Zimmermann and Brox, 2017]
- Wider Face [Yang et al., 2016]
- Multi-Human Parsing [Zhao et al., 2018]

# Experimentos e Resultados

---

# Configuração

GCP Compute Engine VM: 8 vCPUs com 30GB de memória e 1 GPU NVIDIA Tesla K80

## Treinamento

Implementação utilizada: Abdulla [2017]

Data augmentation: espelhamento horizontal, rotação de 90° ou 180° e *Gaussian blurring*

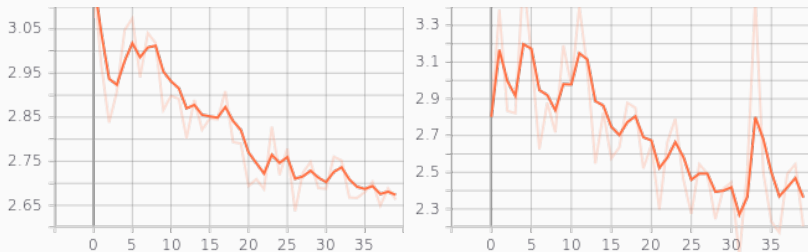
1. 30 épocas + rede pré-treinada no dataset COCO [Lin et al., 2014]
2. 10 épocas de fine tuning a partir da 4ª camada

## Validação

$mAP_{IOU=0.5}$ , 20 rodadas de inferência em 40 imagens rotuladas  
Inspeção visual de 100 quadros selecionados aleatoriamente do Libras-20 [Almeida et al., 2019]

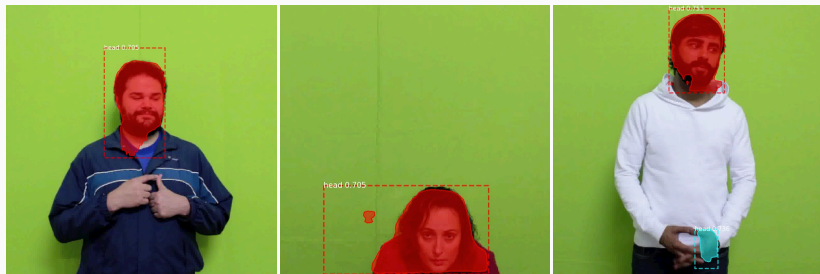
Tempo de treinamento: 13 horas

mAP<sub>IOU=0.5</sub>:  $19.87 \pm 4.56$



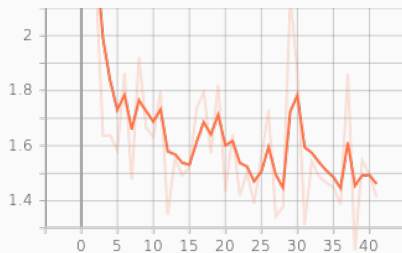


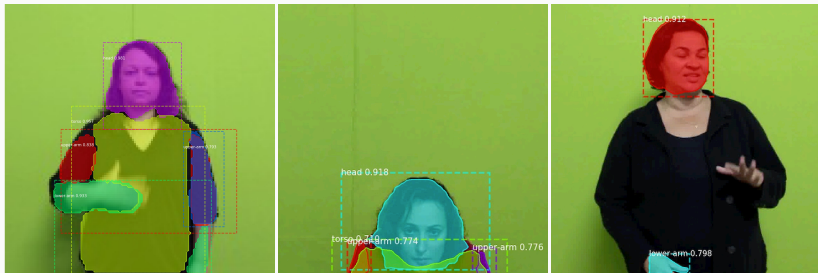
# PASCAL-Person-Part: Libras-20



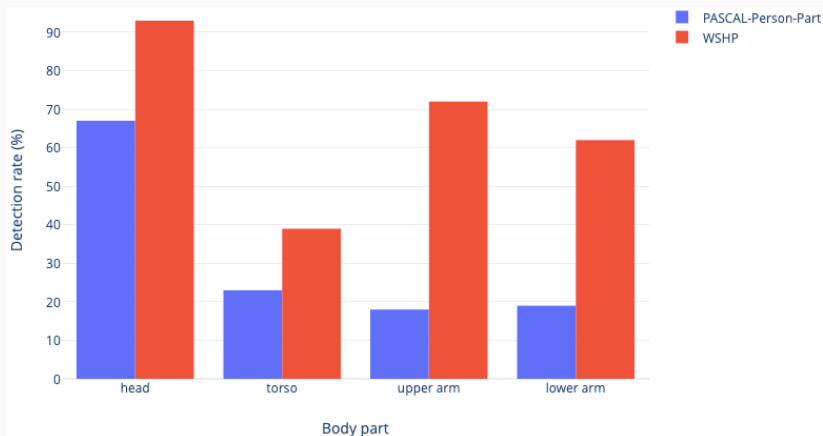
Tempo de treinamento: 52 horas

$mAP_{IOU=0.5}$ :  $32.09 \pm 2.14$





# Libras-20: Resultados de Detecção



## Considerações Finais

---

## **WSHP completo**

Treinamento com todas as imagens disponíveis.

## **Modelos separados**

Treinar um modelo para cada região de interesse.

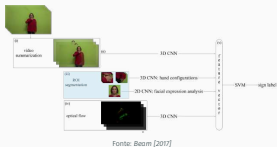
## **Arquiteturas diferentes**

Treinar com arquiteturas mais recentes, como a FastFCN [Wu et al., 2019] ou a Gated Shape CNN [Takikawa et al., 2019].

## **Integração com o SRS**

Adicionar o modelo treinado ao fluxo de reconhecimento de sinais e testar o sistema como um todo.

## Sistema de Reconhecimento de Sinais (SRS)



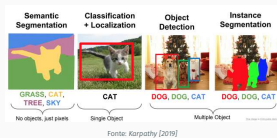
Rúbia Guerra (Escola de Engenharia, UFMG)

Deep Learning for Accessibility

Introdução

6

## Segmentação de Instâncias



Rúbia Guerra (Escola de Engenharia, UFMG)

Deep Learning for Accessibility

Fundamentos

7

## Weakly and Semi Supervised Human Body Part Parsing via Pose-Guided Knowledge Transfer (WSHP)



WSHP: 150 mil imagens  
Fonte: Fang et al. [2018]

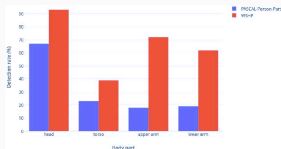
Rúbia Guerra (Escola de Engenharia, UFMG)

Deep Learning for Accessibility

Recursos de Dados

10

## Libras-20: Resultados de Detecção



Rúbia Guerra (Escola de Engenharia, UFMG)

Deep Learning for Accessibility

Experimentos e Resultados

18

Grata pela presença de todos!

<https://rubia-rg.github.io/>  
<https://minds.eng.ufmg.br/>



- W. Abdulla. Mask r-cnn for object detection and instance segmentation on keras and tensorflow.  
[https://github.com/matterport/Mask\\_RCNN](https://github.com/matterport/Mask_RCNN), 2017.
- S. G. M. Almeida, F. G. Guimarães, and J. A. Ramírez. Feature extraction in brazilian sign language recognition based on phonological structure and using rgb-d sensors. *Expert Systems with Applications*, 41(16):7259–7271, 2014.
- S. G. M. Almeida, T. M. Rezende, G. T. B. Almeida, A. C. R. Toffolo, and F. G. Guimarães. Libras-20 dataset, July 2019. URL <https://doi.org/10.5281/zenodo.2667329>.

## Referências ii

- L. Amaral, G. L. N. Júnior, T. Vieira, and T. Vieira. Evaluating deep models for dynamic brazilian sign language recognition. In R. Vera-Rodriguez, J. Fierrez, and A. Morales, editors, *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, page 930–937. Springer International Publishing, 2019. ISBN 978-3-030-13469-3.
- S. Bambach, S. Lee, D. J. Crandall, and C. Yu. Lending a hand: Detecting hands and recognizing activities in complex egocentric interactions. In *The IEEE International Conference on Computer Vision (ICCV)*, December 2015.
- A. L. Beam. Deep learning 101 - part 1: History and background. Online, 2017. URL [https://beamandrew.github.io/deeplearning/2017/02/23/deep\\_learning\\_101\\_part1.html](https://beamandrew.github.io/deeplearning/2017/02/23/deep_learning_101_part1.html). [Accessed June 10th, 2019].

- X. Chen, R. Mottaghi, X. Liu, S. Fidler, R. Urtasun, and A. Yuille. Detect what you can: Detecting and representing objects using holistic models and body parts. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1971–1978, 2014.
- A. Er-Rady, R. Faizi, R. O. H. Thami, and H. Housni. Automatic sign language recognition: A survey. In *2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, page 1–7. IEEE, May 2017. ISBN 978-1-5386-0551-6. URL <http://ieeexplore.ieee.org/document/8075561/>.
- H.-S. Fang, G. Lu, X. Fang, J. Xie, Y.-W. Tai, and C. Lu. Weakly and semi supervised human body part parsing via pose-guided knowledge transfer. *CVPR*, 2018.

- C. F. F. C. Filho, R. S. de Souza, J. R. dos Santos, B. L. dos Santos, and M. G. F. Costa. A fully automatic method for recognizing hand configurations of brazilian sign language. *Research on Biomedical Engineering*, 33(1):78–89, mar 2017. URL <https://doi.org/10.1590/2446-4740.03816>.
- R. Gandhi. R-cnn, fast r-cnn, faster r-cnn, yolo — object detection algorithms, Jul 2018. URL <https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-a> [Accessed June 2nd, 2019].
- HandTalk. Handtalk. Online, 2019. URL <http://www.handtalk.me/>. [Accessed May 31st, 2019].
- G. Hirata. Existem gírias em libras? | superinteressante, 2019. URL <https://super.abril.com.br/cultura/existem-girias-na-lingua-de-sinais-dos-surdos/>.

## Referências v

- A. Karpathy. Cs231n convolutional neural networks for visual recognition, 2019. URL <http://cs231n.github.io/transfer-learning/#tf>.
- A. Kuenburg, P. Fellingner, and J. Fellingner. Health care access among deaf people. *The Journal of Deaf Studies and Deaf Education*, 21(1): 1–10, Jan 2016. ISSN 1081-4159. doi: 10.1093/deafed/env042.
- T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.
- L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikäinen. Deep learning for generic object detection: A survey. *International Journal of Computer Vision*, Oct 2019. ISSN 1573-1405. doi: 10.1007/s11263-019-01247-4. URL <https://doi.org/10.1007/s11263-019-01247-4>.

## Referências vi

- T. M. Rezende. Aplicação de técnicas de inteligência computacional para análise da expressão facial em reconhecimento de sinais de libras. Master's thesis, UFMG, 2016. URL <https://www.ppgee.ufmg.br/defesas/1393M.PDF>.
- T. Takikawa, D. Acuna, V. Jampani, and S. Fidler. Gated-scnn: Gated shape cnns for semantic segmentation. *CoRR*, abs/1907.05740, 2019. URL <http://arxiv.org/abs/1907.05740>.
- VLibras. Vlibras - tradução de português pra libras. Online, 2019. URL <http://www.vlibras.gov.br/>. [Accessed May 31st, 2019].
- L. Weng. Object detection for dummies part 3: R-cnn family, Dec 2017. URL <https://lilianweng.github.io/2017/12/31/object-recognition-for-dummies-part-3.html>.
- WHO. Who | estimates. Online, 2019. URL <http://www.who.int/deafness/estimates/en/>.

## Referências vii

- H. Wu, J. Zhang, K. Huang, K. Liang, and Y. Yizhou. Fastfcn: Rethinking dilated convolution in the backbone for semantic segmentation. In *arXiv preprint arXiv:1903.11816*, 2019.
- S. Yang, P. Luo, C. C. Loy, and X. Tang. Wider face: A face detection benchmark. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- J. Zhao, J. Li, Y. Cheng, L. Zhou, T. Sim, S. Yan, and J. Feng. Understanding humans in crowded scenes: Deep nested adversarial learning and a new benchmark for multi-human parsing. *arXiv preprint arXiv:1804.03287*, 2018.
- C. Zimmermann and T. Brox. Learning to estimate 3d hand pose from single rgb images. Technical report, arXiv:1705.01389, 2017. URL <https://lmb.informatik.uni-freiburg.de/projects/hand3d/>. <https://arxiv.org/abs/1705.01389>.