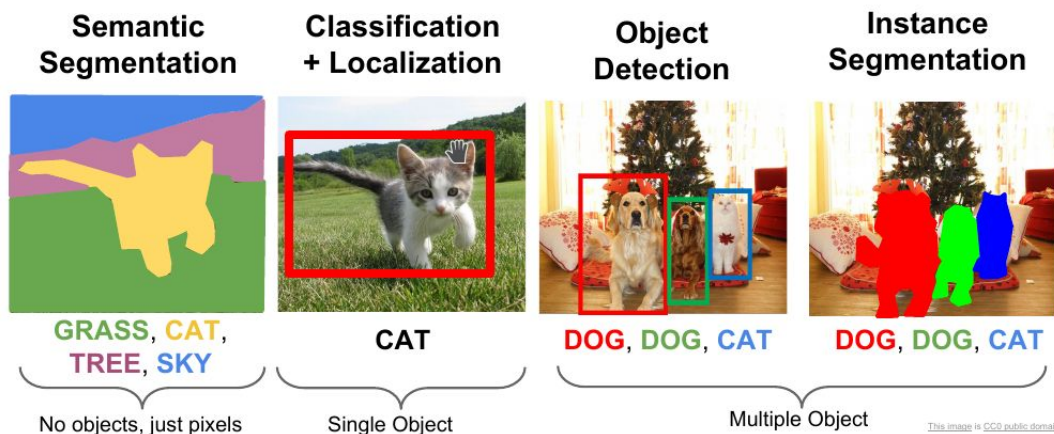


## Applications of deep learning: understanding images

This includes problems such as: locating an object within an image, classifying an object in an image, detecting text within an image, estimating the pose of people within an image.



*Image: Examples of problems within this topic, via [Stanford CS231 slides](#)*

There are many topics in this area, so I have grouped them by problem type.

### Image classification

#### 1a. SketchTransfer

Citation: Alex Lamb Sherjil Ozair Vikas Verma David Ha. SketchTransfer: A Challenging New Task for Exploring Detail-Invariance and the Abstractions Learned by Deep Networks. In Proceedings of the 2020 IEEE Winter Conference on Applications of Computer Vision (WACV '20). [\[PDF\]](#)

Note: this project is about a particular *task*, not a model. The task involves training a model on CIFAR-10 and then testing it on Quick Draw. (More details in the paper.)

#### 1b. One-shot classification with Siamese networks

Citation: Gregory Koch, Richard Zemel, Ruslan Salakhutdinov. Siamese Neural Networks for One-shot Image Recognition. In Proceedings of the Deep Learning Workshop at ICML '15. [\[PDF\]](#) [\[Github\]](#) (unofficial, includes notebook) [\[Github\]](#) (unofficial) [\[Github\]](#) (unofficial)

#### 1c. Few-shot image classification with REPTILE

Citation: Alex Nichol, Joshua Achiam, John Schulman. On First-Order Meta-Learning Algorithms. arXiv:1803.02999 (2018). [\[PDF\]](#) [\[Blog post\]](#) [\[Github\]](#) (official) [\[Tutorial\]](#) on image classification, with notebook [\[Tutorial\]](#) notebook on REPTILE for regression

## Object detection

### 1d. Small and fast object detection: MobileNet + SSD

Citation (SSD): Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg. SSD: Single Shot MultiBox Detector. In Proceedings of the 2016 European Conference on Computer Vision (ECCV '16). [[PDF](#)]

Citation (MobileNet): Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv:1704.04861 (2017). [[PDF](#)] [[Blog post](#)]

[Tensorflow [object detection tutorial](#) (change the module to MobileNetV2+SSD)]  
[[Example](#): fine tuning for blood cell detection] [SSD [demo notebook](#)]

### 1e. YOLO: real time object detection

Citation: Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '16). [[PDF](#)]

(Also see updates: YOLOv2, YOLOv3, YOLOv4) [[Website](#)]

[[Github](#) (includes [notebook for object detection](#), [notebook for training](#))]

[[Notebook](#) (unofficial): Object detection YOLOv3]

### 1f. Object detection with transformers (DETR)

Citation: Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, Sergey Zagoruyko. End-to-End Object Detection with Transformers. arXiv:2005.12872 (2020) [[PDF](#)] [[Blog post](#)]

[[Github](#) (official via Facebook, with notebooks)]

### 1g. One-stage face detection

Citation: Jiankang Deng, Jia Guo, Yuxiang Zhou, Jinke Yu, Irene Kotsia, Stefanos Zafeiriou. RetinaFace: Single-stage Dense Face Localisation in the Wild. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR '20). [[PDF](#)][[Github](#) (official)] [[Tutorial](#) (pre-trained model)]

[[Github](#) (unofficial, includes Colab notebook)]

[More pre-trained face detection models, in a Colab [notebook](#) (use TF 1!)]

### 1h. Multi-task Self-supervised Object Detection

Citation: Wonhee Lee, Joonil Na, Gunhee Kim. Multi-task Self-supervised Object Detection via Recycling of Bounding Box Annotations. In Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '19). [[PDF](#)]

[[Github](#) (official, includes [notebook](#))]

## Object tracking

### 1i. Multiple object tracking

Citation (SORT): Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, Ben Upcroft. Simple Online and Realtime Tracking. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP '16). [[PDF](#)] [[Github](#)] [[Notebook](#)]

Citation (DeepSORT): Nicolai Wojke, Alex Bewley, Dietrich Paulus. Simple Online and Realtime Tracking with a Deep Association Metric. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP '17). [[PDF](#)] [[Github](#)] (official), + [this repo](#)] [[Github](#)] (unofficial)] [[Github](#)] (unofficial)] [[Notebook](#)]

### 1j. Tracking with SiamMASK

Citation: Qiang Wang, Li Zhang, Luca Bertinetto, Weiming Hu, Philip H.S. Torr. Fast Online Object Tracking and Segmentation: A Unifying Approach. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR '19). [[PDF](#)] [[Github](#)] [[Video](#)] [[Website](#)] [[Notebook](#)]

## Semantic segmentation/instance segmentation

### 1k. DeepLab

Citation: Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, Alan L. Yuille. In IEEE Transactions on Pattern Analysis and Machine Intelligence (2017). (Also see later updates). [[PDF](#)] [[Github](#)] - Tensorflow, [notebook](#) for TPU] [DeepLab3 is in [Detectron2](#)]

### 1l. ShapeMask

Citation: Weicheng Kuo, Anelia Angelova, Jitendra Malik, Tsung-Yi Lin. ShapeMask: Learning to Segment Novel Objects by Refining Shape Priors. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV '19). [[PDF](#)] [[Video](#)] [[Notebook](#)] (via tensorflow on GitHub)]

### 1m. Mask R-CNN

Citation: Kaiming He, Georgia Gkioxari, Piotr Dollar, Ross Girshick. Mask R-CNN. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV '17). [[PDF](#)] [[Video](#)] [[Notebook](#)] (via tensorflow on Github)] [[Notebook](#)] (via Facebook Detectron2)] [[Github](#)] Mask R-CNN on Python 3, Keras, and TensorFlow, including notebooks] [More notebooks (via tugstugi on GitHub): [Detectron torchvision](#)]

### 1n. YOLACT

Citation: Daniel Bolya, Chong Zhou, Fanyi Xiao, Yong Jae Lee. YOLACT: Real-time Instance Segmentation. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV '19). [[PDF](#)] [[Video](#)]

Citation: Daniel Bolya, Chong Zhou, Fanyi Xiao, Yong Jae Lee. YOLACT++: Better Real-time Instance Segmentation. arXiv:1912.06218 (2019) [[PDF](#)] [[Github](#)] [[Notebook](#)] [Another [notebook](#)]

## 3D shape understanding

### 1o. Semantic mesh segmentation with graph convolutions

Citation: Nitika Verma, Edmond Boyer, Jakob Verbeek. FeaStNet: Feature-Steered Graph Convolutions for 3D Shape Analysis. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '18). [[PDF](#)] [[Github](#) (official)] [[Notebook](#)] [[PyTorch implementation](#)]

### 1p. High-resolution 3D human digitization

Citation: Shunsuke Saito, Tomas Simon, Jason Saragih, Hanbyul Joo. PIFuHD: Multi-Level Pixel-Aligned Implicit Function for High-Resolution 3D Human Digitization. In Proceedings of the 2020 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '20). [[PDF](#)] [[Website](#)] [[Video](#)] [[Github](#)] [[Notebook](#)]

## Pose estimation

### 1q. DensePose

Citation: Rıza Alp Güler, Natalia Neverova, Iasonas Kokkinos. DensePose: Dense Human Pose Estimation in the Wild. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '18). [[PDF](#)] [[GitHub Repo](#) (Facebook Research)] [[Notebook](#) (via tugstugi on GitHub)]

### 1r. VideoPose3D

Citation: Dario Pavlo, Christoph Feichtenhofer, David Grangier, Michael Auli. 3D Human Pose Estimation in Video With Temporal Convolutions and Semi-Supervised Training. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR '19). [[PDF](#)] [[Overview](#)] [[GitHub Repo](#) (Facebook Research)] [[Notebook](#) (via Justin Emmerich on GitHub)]

### 1s. VIBE: Video Inference for Human Body Pose and Shape Estimation

Citation: VIBE: Video Inference for Human Body Pose and Shape Estimation. Muhammed Kocabas, Nikos Athanasiou, Michael J. Black; Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 5253-5263. [[PDF](#) and [extras](#)] [[Official Github release](#) (includes Colab notebook)]

## Action recognition

### 1t. Action recognition with I3D

Citation: Joao Carreira, Andrew Zisserman. Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '17). [[PDF](#)] [[Notebook](#) (pretrained model)] [Also in gluon: [intro tutorial](#), [training tutorial](#)]

#### 1u. Action recognition with R(2+1)D

Citation: Du Tran, Heng Wang, Lorenzo Torresani, Jamie Ray, Yann LeCun, Manohar Paluri. A Closer Look at Spatiotemporal Convolutions for Action Recognition. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '18). [[PDF](#)]

Citation: Deepti Ghadiyaram, Matt Feiszli, Du Tran, Xueting Yan, Heng Wang, Dhruv Mahajan. Large-scale weakly-supervised pre-training for video action recognition. In Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '19). [[PDF](#)] [[Pre-trained model in gluon](#)] [[Github](#)] (via Microsoft, pretrained model, with notebooks)]

Multimodal applications (involves understanding images *and* text or audio):

#### 1v. Visual question answering

Citation: Aishwarya Agrawal, Jiasen Lu, Stanislaw Antol, Margaret Mitchell, C. Lawrence Zitnick, Dhruv Batra, Devi Parikh. VQA: Visual Question Answering. In Proceedings of the International Conference on Computer Vision (ICCV '15). [[PDF](#)] [[Github](#) PyTorch (unofficial)] [[Github](#) Tensorflow (unofficial, with notebook, [blog](#))] [Easy VQA: a "toy" example. See [Github](#) for dataset, [Github](#) for model, [blog post](#)]

Citation: Yu Jiang, Vivek Natarajan, Xinlei Chen, Marcus Rohrbach, Dhruv Batra, Devi Parikh. Pythia v0. 1: the winning entry to the vqa challenge 2018. arXiv preprint arXiv:1807.09956. 2018. [[PDF](#)]

Citation: Amanpreet Singh, Vivek Natarajan, Meet Shah, Yu Jiang, Xinlei Chen, Dhruv Batra, Devi Parikh, Marcus Rohrbach. Towards VQA Models That Can Read. In Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '19). [[PDF](#)] [[Notebook](#)] [[Github](#)]

#### 1w. Speech reconstruction from silent video

Citation: Ariel Ephrat, Shmuel Peleg. Vid2speech: Speech Reconstruction from Silent Video. In Proceedings of the 2017 International Conference on Acoustics, Speech, and Signal Processing (ICASSP '17). [[PDF](#)] [[Github](#)]

Citation: Ariel Ephrat, Tavi Halperin, Shmuel Peleg. Improved Speech Reconstruction from Silent Video. In Proceedings of the 2017 Workshop on Computer Vision for Audio-Visual Media (ICCV '17). [[PDF](#)] [[Website](#)]

Citation: K R Prajwal, Rudrabha Mukhopadhyay, Vinay Namboodiri, C V Jawahar. Learning Individual Speaking Styles for Accurate Lip to Speech Synthesis. In Proceedings of the 2020 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '20). [[PDF](#)] [[Github](#)] [[Website](#)] [[Video](#)]