

Applied AI



brought you by:
inzva, Bev Foundation

Guides:

Halil Eralp Koçaş

Ahmet Melek

Gürkan Soykan

Şafak Bilici & Enes Sadi Uysal

Abdurrahman Dilmaç

Duygu Ay

This Week: Computer Vision

- Five Problems:

- Classification +
- Localization +
- Semantic Segmentation +
- Object Detection +
- Instance Segmentation +
- Image Generation

- Three model types:

- Dense (Fully-Connected) Networks +
- Convolutional Neural Networks +
- Siamese Neural Networks +
- Autoencoders
- Generative Adversarial Networks

- Four Frameworks

- Tensorflow 2
- Keras Functional +
- Keras Sequential +
- PyTorch & torchvision +

- Six Datasets:

- MNIST +
- CIFAR-10/100 +
- Kaggle Facial Keypoint Detection +
- Segmentation +
- COCO for Object Detection +
- Occlusion Dataset

Today's Schedule

- Introduction 10.00 - 10.15
- Similarity Learning on MNIST with Siamese Conv Networks on Keras Functional 10.15 - 11.00
- **Break** **11.00 - 11.15**
- Kaggle Facial Keypoints Detection with Conv Nets on PyTorch 11.15 - 12.00
- **Lunch Break** **12.00 - 12.45**
- CIFAR-10 Classification and Transfer Learning with ConvNets on Keras Sequential 12.45 - 13.30
- Semantic Segmentation with U-Net Architecture 13.30 - 14.15
- **Break** **14.15 - 14.30**
- Object Detection with a Pre-Trained model from Torchvision 14.30 - 14.50
- Instance Segmentation with a Pre-Trained model from Torchvision 14.50 - 15.10
- Homework Description 15.10 - 15.30

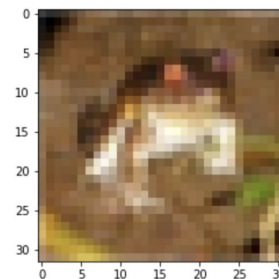
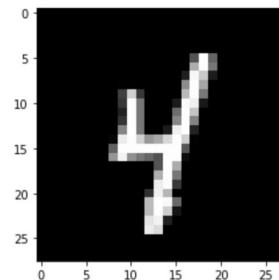
Computer Vision



- Humans can retrieve basic information and many more from looking an image.
 - Four people, many cars, street, etc.
 - People are walking, one of them is barefoot, etc.
 - Their clothes, danger of getting hit by a car, etc.
- Thus, we aim to reach at least this level using Computer Vision algorithms.
- A system must recognize these examples we talked above and more to provide a description as complete as possible of the image.

Classification

- Classification of a given image as belonging to one of a set of predefined categories.
- MNIST Dataset:
 - Handwritten digit classification
 - 60000 training, 10000 test examples
 - Supervised problem
- CIFAR-10 Dataset:
 - Image classification
 - 50000 training, 10000 test examples
 - 32x32x3 images, supervised problem
 - 10 classes, 5000 training, 1000 test each
- CIFAR-100 Dataset:
 - Same settings as CIFAR-10
 - 500 training and 100 test per class
 - 100 classes



Similarity Learning with Siamese Neural Networks

- Goal: understand whether two inputs are same or not.
 - Understanding **similarity**!
- Classification for pre-defined number of classes does not scale.
- Siamese NNs:
 - Why calling Siamese?
 - Same weights are shared for inputs.
 - Process two images in parallel
 - Decide based on a distance function
 - Loss functions:
 - Classification losses work
 - Binary cross-entropy
 - Triplet loss

Keras Functional

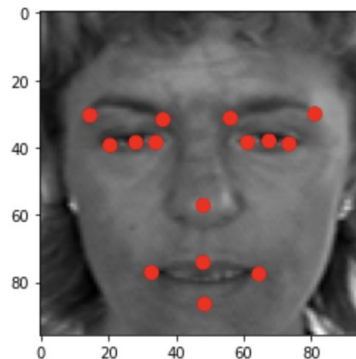
- Create more flexible models than Keras Sequential!
 - Functional API can handle models with non-linear topology, shared layers, and multiple inputs & outputs.
- DL models -> Directed Acyclic Graphs (DAG). Functional is a way to build graphs of layers.
- Training, evaluation, inference, and saving models are exactly same for both Functional and Sequential API.
- Can use same layers for multiple models!

Keras Sequential

- Use when a model is appropriate for **a plain stack of layers** where each layer has **exactly one input tensor and one output tensor**.
- Once a Sequential model is built, it behaves like Functional API model.
 - Each layer has input and output.
 - Can create a new model to observe the output of each layer, etc.

Localization

- Goal is to learn where the objects are in given images.
- Standard way to identify localization in images using bounding boxes to encapsulate localized objects.
- Kaggle Facial Keypoints Dataset
 - Predict keypoint positions on face images.
 - 7049 training, 1783 test images
 - 96x96x1 images
 - Supervised problem
 - 2D 15 features

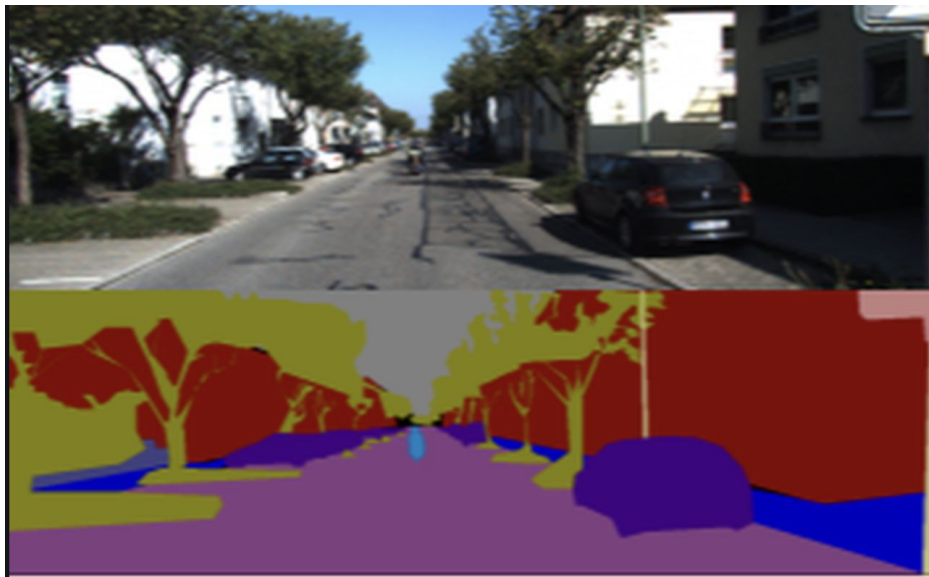


PyTorch

- Optimized tensor library for deep learning using GPUs and CPUs.
- Many libraries are part of PyTorch project for many tasks such as audio, text, vision, etc.
- Ease-of-use and flexible.

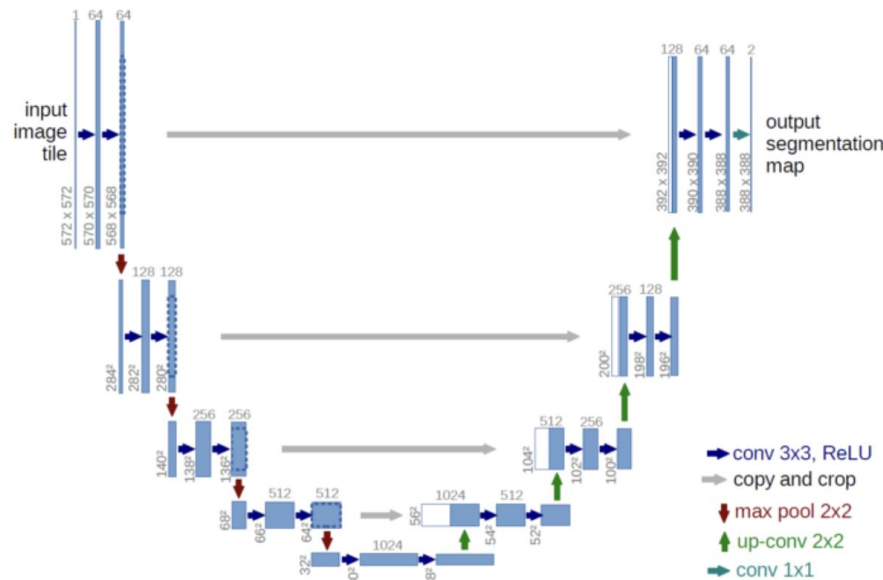
Semantic Segmentation

- Goal: predict class of each pixel in image.
- Output: generally a high resolution output image with same shape of input image with each class is represented with a color.



U-Net Architecture

- Encoder-decoder architecture with skip-connections.
 - Conv layers
 - Max pool layers
 - Up-conv layers
 - Skip-connection layers
- Bottleneck layer for high-level information.
- Skip-connections are for low-level information.
 - Where you utilize depends on use case.

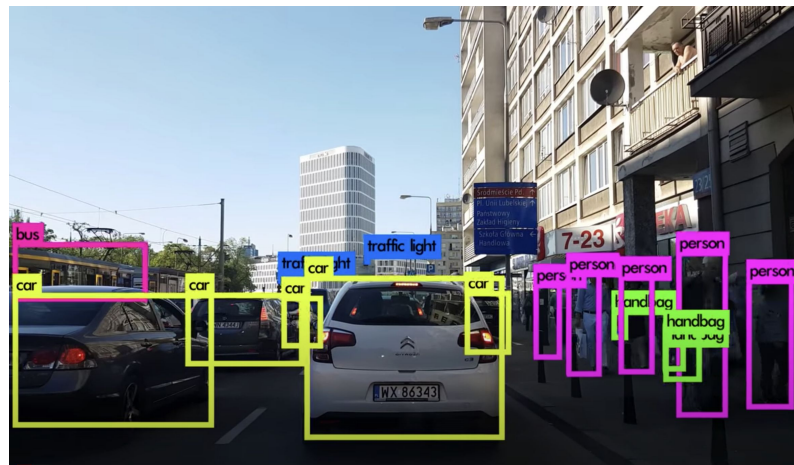


torchvision

- Part of PyTorch project.
- The torchvision package consists of popular datasets, model architectures, and common image transformations for computer vision.

Object Detection

- Localization and classification of all objects present in a given image.
- Can be multiple classes.
- Bounding boxes to localize and classification to understand what object is in the bounding box.
- Complex Problem
 - Some objects are occluded, partially visible, etc.
- COCO Dataset
 - There are images for all tasks
 - Object detection, segmentation, etc.
 - 330k images (>200k labeled)
 - 80 object, 91 stuff categories



Instance Segmentation

- Goal: treats each objects of the same class as they are different objects.
- Output: generally a high resolution output image with same shape of input image with each object is represented with a different color.

