Deep Learning with Computer Vision

SEMINAR

contents

mobile:+91.9880951838
mailto:mohit.riverstone@gmail.com
mailto:mohit@stillwaters.ai
website:www.stillwaters.ai
blog:slowbreathing.github.io

Deep Learning with Computer Vision

SEMINAR

course contents

- Software Tools Day1
- Convolutional Neural Networks Internals
- Deep Generative Models Day2
- OpenCV
- Deep learning for computer vision Day3

mobile:+91.9880951838
mailto:mohit.riverstone@gmail.com
mailto:mohit@stillwaters.ai
website:www.stillwaters.ai
blog:slowbreathing.github.io

Description:

- You’ve probably heard that Deep Learning is making news across the world as one of the most promising techniques in machine learning, especially for analyzing image data. With every industry dedicating resources to unlock the deep learning potential, to be competitive, you will want to use these models in tasks such as image tagging, object recognition, speech recognition, and text analysis. In this training session you will build deep learning models for Computer Vision. One the detailed case study will be using attention based mechanism to do image segmentation and object recognition. Another detailed case study would be image captioning and understanding using a combination attention based CNN and sequence based model(LSTM).

Key Skills:

- face recognition
- image generation
- video classification
- image captioning
- medical image segmentation
- product search
- Optical Character/Word/Sentence Recognition
- person detection

Prerequisites:

- This is an advanced level session and it assumes that you have good familiarity with Machine learning.
- Working Knowledge of python
- Machine Learning Internals
- All sequence based models like RNN, LSTM, GRUs, Attention,Language Models must be known as in "Modern Natural Language Processing(NLP) with Deep Learning session(Kindly refer to the first item in the brief TOC)".

Instructional Method:

- This is an instructor led course provides lecture topics and the practical application of Deep Learning and the underlying technologies. It pictorially presents most concepts and there is a detailed case study that strings together the technologies, patterns and design.

Deep Learning with Computer Vision

- ***Software Tools***
  - Tensorflow
    - Installation
    - Sharing Variables
    - Creating Your First Graph and Running It in a Session
    - Managing Graphs
    - Visualizing the Graph and Training Curves Using TensorBoard
    - Implementing Gradient Descent
    - Lifecycle of a Node Value
    - Linear Regression with TensorFlow
    - Modularity
    - Saving and Restoring Models
    - Name Scopes
    - Feeding Data to the Training Algorithm
  - Keras
- ***Convolutional Neural Networks Internals***
  - Pooling layer
  - Image augmentation
  - Convolutional layer
  - History of CNNs
  - Convolutional layers in Keras
  - Code for visualizing an image
  - Input layer
  - How do computers interpret images?
  - Practical example image classification
  - Convolutional neural networks
  - Dropout
  - Attention Mechanism for CNN and Visual Models
    - Types of Attention
    - Glimpse Sensor in code

  - Attention mechanism for image captioning
  - Hard Attention
  - Applying the RAM on a noisy MNIST sample
  - Recurrent models of visual attention
  - Using attention to improve visual models
  - Reasons for sub-optimal performance of visual CNN models
  - Soft Attention

- Build Your First CNN and Performance Optimization
  - Convolution and pooling operations in TensorFlow
  - Convolutional operations
  - Using tanh
  - Convolution operations in TensorFlow
  - Regularization
  - Fully connected layer
  - Weight and bias initialization
  - Pooling, stride, and padding operations
  - CNN architectures and drawbacks of DNNs
  - Applying pooling operations in TensorFlow
  - Using sigmoid
  - Training a CNN
  - Using ReLU
  - Activation functions
  - Building, training, and evaluating our first CNN
    - Creating a CNN model
    - Defining CNN hyperparameters
    - Model evaluation
    - Dataset description
    - Loading the required packages
    - Running the TensorFlow graph to train the CNN model
    - Preparing the TensorFlow graph
    - Loading the training/test images to generate train/test set
    - Constructing the CNN layers

- Model performance optimization
    - Applying dropout operations with TensorFlow
    - Building the second CNN by putting everything together
    - Appropriate layer placement
    - Which optimizer to use?
    - Creating the CNN model
    - Dataset description and preprocessing
    - Number of neurons per hidden layer
    - Number of hidden layers
    - Batch normalization
    - Memory tuning
    - Training and evaluating the network
    - Advanced regularization and avoiding overfitting

- Popular CNN Model Architectures
    - Architecture insights
    - ResNet architecture
    - AlexNet architecture
    - VGG image classification code example
    - Introduction to ImageNet
    - VGGNet architecture
    - GoogLeNet architecture
    - LeNet
    - Traffic sign classifiers using AlexNet
    - Inception module

- Transfer Learning
    - Multi-task learning
    - Target dataset is small but different from the original training dataset
    - Autoencoders for CNN
    - Applications
    - Target dataset is large and similar to the original training dataset
    - Introducing to autoencoders

  - Convolutional autoencoder
  - Target dataset is large and different from the original training dataset
  - Transfer learning example
  - Feature extraction approach
  - Target dataset is small and is similar to the original training dataset
  - An example of compression

- GAN: Generating New Images with CNN
  - Feature matching
  - GAN code example
  - Deep convolutional GAN
  - Adding the optimizer
  - Training a GAN model
  - Semi-supervised learning and GAN
  - Pixpix - Image-to-Image translation GAN
  - Calculating loss
  - Semi-supervised classification using a GAN example
  - CycleGAN
  - Batch normalization
- Object Detection and Instance Segmentation with CNN
  - Creating the environment
  - Fast R-CNN (fast region-based CNN)
  - The differences between object detection and image classification
  - Mask R-CNN (Instance segmentation with CNN)
  - Cascading classifiers
  - Haar Features
  - Faster R-CNN (faster region proposal network-based CNN)
  - Traditional, nonCNN approaches to object detection
  - R-CNN (Regions with CNN features)
  - Running the pre-trained model on the COCO dataset
  - Why is object detection much more challenging than image

classification?

  - The Viola-Jones algorithm
  - Preparing the COCO dataset folder structure
  - Downloading and installing the COCO API and detectron library (OS shell commands)
  - Instance segmentation in code
  - Haar features, cascading classifiers, and the Viola-Jones algorithm
  - Installing Python dependencies (Python environment)

- Popular CNN Model Architectures
  - Introduction to ImageNet
  - VGG image classification code example
  - GoogLeNet architecture
  - Architecture insights
  - Inception module
  - AlexNet architecture
  - VGGNet architecture
  - LeNet
  - ResNet architecture
  - Traffic sign classifiers using AlexNet

- ***Deep Generative Models***
  - Deep Boltzmann Machines
  - Back-Propagation through Random Operations
  - Restricted Boltzmann Machines
  - Generative Stochastic Networks
  - Boltzmann Machines for Structured or Sequential Outputs
  - Boltzmann Machines
  - Other Boltzmann Machines
  - Other Generation Schemes
  - Directed Generative Nets
  - Boltzmann Machines for Real-Valued Data
  - Evaluating Generative Models
  - Drawing Samples from Autoencoders

  - Deep Belief Networks
  - Convolutional Boltzmann Machines
- ***OpenCV***
  - The Core Functionality (core module)
  - Introduction to OpenCV
  - Object Detection (objdetect module)
  - Image Processing (imgproc module)
  - Deep Neural Networks (dnn module)
  - GPU-Accelerated Computer Vision (cuda module)
- ***Deep learning for computer vision***
  - Similarity learning
    - Human face analysis
    - Face landmarks and attributes
      - Multi-Task Facial Landmark (MTFL) dataset
      - The Kaggle keypoint dataset
      - The Multi-Attribute Facial Landmark (MAFL) dataset
      - Learning the facial key points
    - Face recognition
      - Finding the optimum threshold
      - The YouTube faces dataset
      - The labeled faces in the wild (LFW) dataset
      - The CelebFaces Attributes dataset
      - CASIA web face database
      - The VGGFace2 dataset
      - Computing the similarity between faces
    - Face detection
    - Face clustering
    - Algorithms for similarity learning
      - Visual recommendation systems
      - DeepRank
      - FaceNet

        - The DeepNet model
        - Contrastive loss
        - Triplet loss
        - Siamese networks
- Classification
    - Image Classification
        - The bigger deep learning models
            - The DenseNet model
            - The Google Inception-V3 model
            - The VGG-16 model
            - The SqueezeNet model
            - The AlexNet model
            - Spatial transformer networks
            - The Microsoft ResNet-50 model
        - Other popular image testing datasets
            - The Fashion-MNIST dataset
            - The CIFAR dataset
            - The ImageNet dataset and competition
        - Training the MNIST model in TensorFlow
            - The MNIST datasets
            - Building a multilayer convolutional network
            - Building a perceptron
            - Loading the MNIST data
        - Training a model for binary classification
            - Transfer learning or fine-tuning of a model
            - Preparing the data
            - Augmenting the dataset
            - Benchmarking with simple CNN
            - Fine-tuning several layers in deep learning
        - Developing real-world applications
            - Brand safety

      - Tackling the underfitting and overfitting scenarios
      - Gender and age detection from face
      - Choosing the right model
      - Fine-tuning apparel models
  - Image Retrieval
    - Model inference
      - Serving the trained model
      - Exporting a model
    - Understanding visual features
      - Embedding visualization
      - The DeepDream
      - Visualizing activation of deep learning models
      - Adversarial examples
      - Guided backpropagation
    - Content-based image retrieval
      - Matching faster using approximate nearest neighbour
      - Extracting bottleneck features for an image
      - Computing similarity between query image and target database
      - Autoencoders of raw images
      - Building the retrieval pipeline
      - Efficient retrieval
      - Advantages of ANNOY
      - Denoising using autoencoders
- Generative models
  - Generative Adversarial Networks
    - Drawbacks of GAN
    - Image translation
    - InfoGAN
    - Conditional GAN
    - Adversarial loss

        - Vanilla GAN
    - Applications of generative models
        - Inpainting
        - Super-resolution of images
        - Blending
        - 3D models from photos
        - Text to image generation
        - Transforming attributes
        - Creating training data
        - Image to image translation
        - Interactive image generation
        - Artistic style transfer
        - Creating new animation characters
        - Predicting the next frame in a video
    - Neural artistic style transfer
        - Style transfer
        - Style loss using the Gram matrix
        - Content loss
    - Visual dialogue model
        - Algorithm for VDM
        - Discriminator
        - Generator
- Video analysis
    - Extending image-based approaches to videos
        - Captioning videos
        - Regressing the human pose
        - Generating videos
        - Tracking facial landmarks
        - Segmenting videos
    - Exploring video classification datasets
        - UCF101

        - YouTube-8M
        - Other datasets
    - Understanding and classifying videos
    - Approaches for classifying videos
        - Using trajectory for classification
        - Multi-modal fusion
        - Using 3D convolution for temporal learning
        - Classifying videos over long periods
        - Fusing parallel CNN for video classification
        - Attending regions for classification
        - Streaming two CNN's for action recognition
- Image captioning
    - Understanding natural language processing for image captioning
        - Expressing words in vector form
        - Training an embedding
        - Converting words to vectors
    - Implementing attention-based image captioning
    - Approaches for image captioning and related problems
        - Retrieving captions from images and images from captions
        - Creating captions using image ranking
        - Using attention network for captioning
        - Using multimodal metric space
        - Knowing when to look
        - Using a condition random field for linking image and text
        - Using RNN on CNN features to generate captions
        - Using RNN for captioning
        - Dense captioning
    - Understanding the problem and datasets
- Detection or localization and segmentation
    - Object Detection

  - Object detection API
    - Re-training object detection models
    - Data preparation for the Pet dataset
    - The YOLO object detection algorithm
    - Monitoring loss and accuracy using TensorBoard
    - Pre-trained models
    - Training the model
    - Object detection training pipeline
    - Training a pedestrian detection for a self-driving car
  - Detecting objects in an image
  - Localizing algorithms
    - Convolution implementation of sliding window
    - Combining regression with the sliding window
    - Thinking about localization as a regression problem
    - Applying regression to other problems
    - The scale-space concept
    - Localizing objects using sliding windows
    - Training a fully connected layer as a convolution layer
  - Detecting objects
    - Single shot multi-box detector
    - Regions of the convolutional neural network (R-CNN)
    - Fast R-CNN
    - Faster R-CNN
  - Exploring the datasets
    - Intersection over Union
    - ImageNet dataset
    - PASCAL VOC challenge
    - COCO object detection challenge
    - Evaluating datasets using metrics
    - The mean average precision
- Semantic Segmentation

- Segmenting satellite images
  - Modeling FCN for segmentation
- Datasets
- Predicting pixels
  - Understanding the earth from satellite imagery
  - Diagnosing medical images
  - Enabling robots to see
- Algorithms for semantic segmentation
  - Large kernel matters
  - The Fully Convolutional Network
  - RefiNet
  - Upsampling the layers by pooling
  - The SegNet architecture
  - DeepLab
  - PSPnet
  - Skipping connections for better training
  - Sampling the layers by convolution
  - Dilated convolutions
- Ultra-nerve segmentation
- Segmenting instances