THIRD EYE

Vision with Deep Learning

Pranab Ghosh

Agenda

1. Deep Learning basics
2. Vision Problems and Solutions
3. Transfer Learning and Pre-trained Models
4. Future Directions

Deep Learning

Deep learning is a very powerful tool that solves many practical problems. However it should never be associated with human level intelligence or any other cognitive skills.

Deep learning models essentially perform statistical correlation in very high dimensional space operating on massive amount of data.

Deep Learning Terminology

- **MLP** : Multi Layer Perceptron. Feed forward network
- **CNN** : Convolution Neural Network. Operates on grid data like image
- **RCNN** : Region based Convolution Neural Network
- **RNN** : Recurrent Neural Network. Operates on sequential data like language
- **LSTM**: Long Short Term Memory. Operates on sequential data
- **GAN** : Generative Adversarial Network. Used to generate data
- **Transformer** : Operates on sequential or grid data. State of the art in DL

Deep Learning Concepts

- Basic elements of a network id neural units and connections and there are multiple layers of it
- The inspiration comes from neural architecture of brain. But the neural circuits in brain is far more complex. It has not been understood completely
- The input is fed in one end of the network and the final output is from the final layer. Output of one layer goes as input to the next layer
- The network weights linearly transform the output from the previous layer to generate input for the current layer.

Deep Learning Concepts

- The input to the current optionally goes through a non linear transformation through activation function before it becomes the final output of the current layer
- A network with one or more hidden layer can approximate any function. That's why a neural network is called called an universal function approximator.
- The network weights are trained to meet some desired task objective
- The error from the final output i.e the difference between actual output and the desired output are back propagated through the network and the network weights adjusted.

Deep Learning Tasks

- In Supervised Learning, the task could be predictive e.g predicting the object in an image. In this case for vision problems training data will consist of images with labels. e.g MLP, RNN, LSTM, CNN
- In Unsupervised Learning, we learn some features or lower dimensional representation which is used for other purposes e.g Auto Encoder.
- In Self Supervised Learning there is no explicit labeling. part of the data is used as label and the model is trained to learn that. In language tasks, some words may be masked and the model is trained to learn the masked words e.g Transformer

Deep Learning Vision Tasks

- Image classification
- Multiple object detection in image
- Image segmentation
- Image generation
- Image caption generation
- Image generation from text
- Image anomaly detection

Object Detection

- The goal is to classify an image i.e type of object it contains. The CNN contains multiple convolution layers followed by a dense network for the final classification. As you go through convolution it extract more high level image features.
- In convolution operation a filter is applied to a patch of an image to produce the feature values. Non linear activation (typically RELU) is applied to the convolution output. This output goose through a pooling layer for down sampling.
- The convolution filter weights are learned through training. Typically multiple filters are used to learn different aspects of the features

Multiple Object Detection

- The goal is to detect multiple objects in an image along with their bounding boxes
- All objects with bounding box are annotated for training. The model predicts the bounding boxes and the objects within
- For each bounding box extracted, CNN based classification is applied. Classification task is more fine grained compared to single object classification in CNN
- Popular models are RCNN, FRCNN and YOLO

Image Segmentation

- The goal is for each pixel to predict what object itl belongs to. It's the most fine grained classification task.
- The network consists of an encoder and decoder. So far we have seen encoder to extract high level features followed by a classifier. The decoder here maps the encoder output to object outlines in the pixel space
- There are different kinds of segmentation. In semantic segmentation, pixels are classified into classes. Instance segmentation classifies pixels into instances of objects. Panoptic segmentation is combination of both.
- Network models are RCNN and CNN

nfo with Link

Image Generation

- The goal of these tasks is opposite of what we have discussed so far. Instead of predictive, the task is generative
- The network used is Generative Adversarial Network(GAN). it consists of a generator that generates realistic image with random noise as input. The discriminator i.e a classifier discriminates between real image and fake images. The discriminator takes real images as input
- As training goes on, the generator learns to generate more realistic images and the discriminator learns to discriminate better between real and fake. Training based on game theory

Cross Modal Generation

- These are also generative tasks but across different modalities i.e text and image
- In image captioning the goal is to train a model so that given an image it can generate a caption as text
- The image captioning architecture uses an encoder decoder pattern. Networks used are CNN, RNN and LSTM
- The reverse task is generating an image from textual description.
- One solution uses GAN. The generator input is text embedding along with some noise and the discriminator is text, image pair

Image Anomaly Detection

- In these taks the goal is to detect anything unusual in an image. Example use cases are medical and manufacturing defect detection
- One solution is based on Auto Encoder based on CNN encoder and decoder. if the image is anomalous, there will be large reconstruction error
- GAN also has been used for anomaly detection. based on the difference between the image and GAN generated images

Transfer Learning and Pre-trained Model

- It’s expensive to train vision models from scratch
- Trained models based millions of publicly available images are available
- You have to fine tune these models for domain specific tasks. It means you train the last few layers of the model while freezing all other layers of the model

Human Visual Processing

- Although CNN is inspired by how visual perception data is processed by out brain, the processing by our brain is far more complex
- According NeuroScience, human visual processing is bi directional. Just like DL, there is visual perception signal flowing bottom up. However there is top down generative signal based what on our past experience and what to expect.
- Our neural circuits learn from the error i.e. difference between the 2 signals

References

- Image Classification
  https://developers.google.com/machine-learning/practica/image-classification/convolutional-neural-networks

- Multi object detection in image:
  https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e

- Image segmentation:
  https://www.v7labs.com/blog/image-segmentation-guide

References

- Image generation
  https://towardsdatascience.com/image-generation-in-10-minutes-with-generative-adversarial-networks-c2afc56bfa3b

- Image caption generation :
  https://towardsdatascience.com/a-guide-to-image-captioning-e9fd5517f350

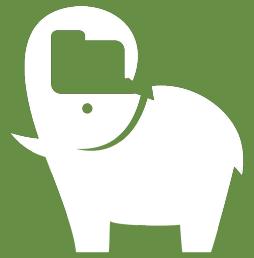
- Image generation from text:
  https://towardsdatascience.com/generating-synthetic-images-from-textual-description-using-gans-e5963bae0df4

Contact

Dj Das

Founder & CEO

djdas@thirdeyedata.io | 408-431-1487 | @djdas



THIRD EYE

Corporate Site	- **ThirdEyeData.io**
Safera Crime Analytics & Predictions	- **Safera.world**
ClouDhiti AI Apps	- **ClouDhiti.ai**
Syra AI Chatbots	- **Syra.ai**
Big Data Cloud Community	- **meetup.com/BigDataCloud**
Phone	- **(408) 462-5257**
Email	- **answers@thirdeyedata.io / consult@thirdeyedata.io**
Twitter	- **@thirdeye_data**
LinkedIn	- **linkedin.com/company/ThirdEyeData**
Facebook	- **facebook.com/ThirdEyeData**
YouTube	- **youtube.com/user/ThirdEyeCSS**
Vimeo	- **vimeo.com/channels/ThirdEyeData**