Kolby D. Boyd

ITAI 1378

SPRING 2024

Table of Contents

Introduction

This course, Introduction to Computer Vision is an overview of how machines can interpret and understand visual data. This course goes into algorithms, techniques, and shows us real-world applications such as autonomous vehicles and medical imaging.

I expect that this class will open my eyes into understanding how computer vision works and how it is already being used in my day-to-day life. I don’t expect the journey to be easy, but I believe Professor McManus is a great teacher and always has an open door if I ever get stuck.

Intro to Computer Vision

Module 01

This module introduced us to our classmates and gave an overview of computer vision. We learned what has been happening in the world of AI and computer vision, such as DALL-, a popular text-to-image generator.

Module 01: Intro to Computer Vision

Main Learnings:

- We can enable machines to understand and interpret **visual** information like humans do.
- Computer Vision is a field of computer science.
- AI is the backbone of this process

Key Results:

- Our group did a powerpoint of a historical timeline of Computer Vision.
- We learned that technology has been a huge motivation due to various worldwide events, such as 9/11 and the need for increased security.

Module 01: Activities & Results

Our group wrote a paper on Real World Applications of Computer Vision. Computer Vision is easily accessible to the masses now, thanks to ChatGPT and DALL-E, among others. Today, we have text-to-video generators as well.

ITAI 1378 – Computer Vision Artificial Intelligence

Andrew Badzioch
Kolby Boyd
Florentin Degbo
Natalia Solozano
Jeralynn Sparks
Professor Patricia McManus

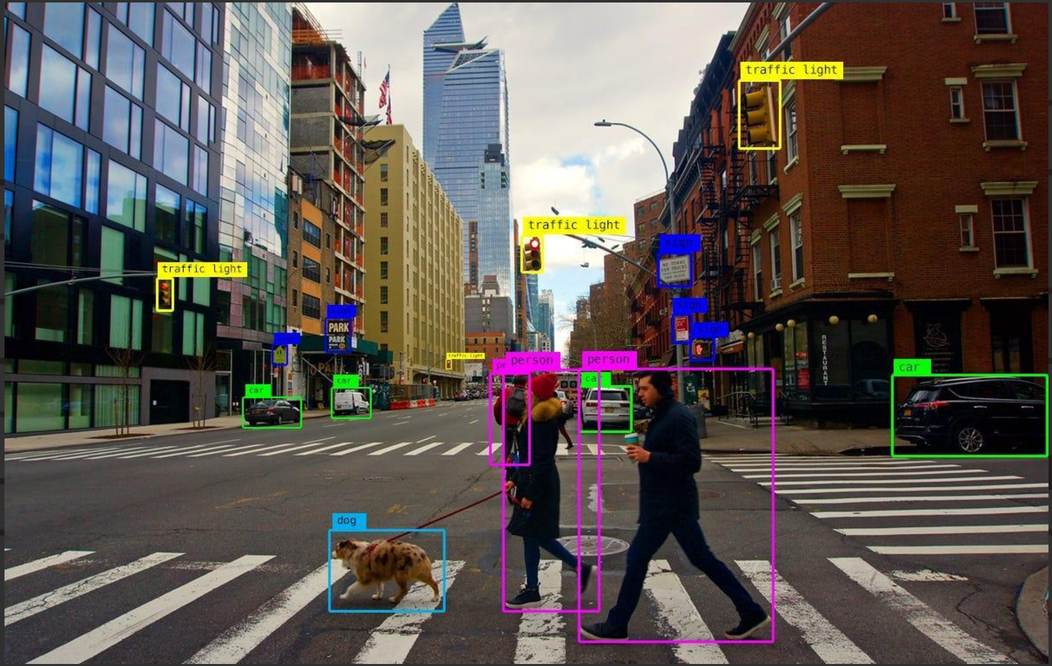
Spring 2024 CN: 15259

L01: Exploring Real-World Applications of Computer Vision

Artificial Intelligence has always been a topic of discussion for many decades now, mostly thought of as a tool to ease the workload of humanity on their day-to-day tasks in the workforce. However, the development of technology has made this tool more readily available, resulting in more innovative and easier ways to let the rest of the population interact with A.I (Artificial Intelligence). on a personal or professional level. This progress brings us to an application called Dall-E 2. Dall-E 2 is an A.I. tool that utilizes different methods, which are discussed later in this essay, to generate images from textual descriptions that are unique in their creation. It also could edit existing photographs through a feature called "in-painting."

One vital question would be: how are these images being created by the application? Dall-E 2 makes use of neural networks which register, code, decode and learn from the datasets provided to it. Beyond just the comprehension of what the text says, it can figure out the relationship between objects and distinguish the

Module 01: Graphs, Tables, & Images



This is an example of what a computer might see through cameras and sensors, and the data the machine is able to glean and remit back to us for further study.

Module 01: References

1. *5 use cases of dalle-3*. KDnuggets. (n.d.). https://www.kdnuggets.com/5-use-cases-of-dalle-3
2. Kerner, S. M. (2023, April 21). *What is dall-e (dall-e 2) and how does it work?*. Enterprise AI. https://www.techtarget.com/searchenterpriseai/definition/Dall-E
3. *Dall·E 2 explained - model architecture, results and comparison*. YouTube. (2022, May 10). https://youtu.be/Z8E3LxqE49M?si=VhtOrPMptBlBZVd9

Intro to Computer Vision #2

Module 02

This module introduced us to the understanding of digital images (pixels, color models, and image formats). We learned about Image processing techniques, such as filtering and edge detection.

Module 02: Intro to Computer Vision #2

Main Learnings:

- Computers see in binary code.
- The learning process is based on a simple neural network:
  - Inputs
  - Hidden layers to detect edges
  - More hidden layers learn larger features
  - Output layer is the classifier
- Prior knowledge (testing) gives the machine a leg up in the process of understanding image data.

Key Takeaway:

- “Image understanding is The process of automating visual tasks by computers, bridging the gap between image processing methods and real-life applications.”

Module 02: Activities & Results

This module had us learn about GitHub and Jupyter Notebooks. I set up my GitHub account, and you can view it here:

https://github.com/kolbyboyd

Kolby Boyd
Prof. Patricia McManus
ITAI 1378
L02

GitHub x Jupyter Notebook

The Setup:

1. **The initial setup of my GitHub account** was straightforward, and I noticed that there were many features that were available to me. I could choose a paid or free plan, and personalize my account to better fit my needs. Since I'm not sure what I will be using GitHub for just yet, I skipped personalization and got straight to work on this project.
   a. Some issues I faced: I was unsure where to start when greeted with my GitHub dashboard. It was all new to me at first, but I closed the ads at the top and found the modules to create my first repository and a README within that repository.
2. **Creating my first repository in GitHub** was a simple process. I was guided by the process and successfully created a repository entitled, *jupyter-exploration* and began to practice making edits and commits to the main branch.
   a. There were no issues during this step of the process.
3. **Installing the Jupyter Notebook** was also simple, but it took some time for the Anaconda installer to complete, although I could not find the Anaconda Navigator on my desktop or in the Start Menu, and when attempting to open from the command prompt or from navigating to the containing folder, the Navigator briefly flashed and immediately closed.
   a. Due to this issue, I went to this website (https://docs.anaconda.com/free/anaconda/install/verify-install/) to verify that my installation was completed correctly. I ended up searching my Local Disc C: to find the folder and created a shortcut on my desktop and taskbar. Then I did some research on GitHub and StackOverflow to discover if anyone else had the same or similar problems which they did. I discovered an article (https://pythoninoffice.com/how-to-fix-anaconda-doesnt-launch/) that described the problem and provided a workable fix by uninstalling PYWIN32 via the command prompt. This didn't work so I uninstalled and reinstalled the Anaconda Navigator. After much troubleshooting, we finally can see the Navigator dashboard in all its glory.
   b. I also needed to install Jupyter Notebook via the Command Prompt for Anaconda to be able to open it.
4. **Performing Basic Operations**
   a. Markdown Cells are cells that hold text. These cells use a language similar to Hypertext Markup Language (HTML), so many of the same functions can be used.
   b. Code Cells are cells that house code, in this case python. When the notebook is run, the code will generate its desired function.
   c. I've learned some basic Markdown Syntax as can be viewed in the Jupyter Notebook.

What I Learned:

1. Concepts
   a. Markdown Language is a very similar, yet lightweight version of HTML (Markup Language).
   b. Combining GitHub and Jupyter Notebooks is unparalleled in its ease of use in collaboration.
2. Tools
   a. GitHub is a great place to store public and private projects for others to collaborate and view.
   b. Jupyter Notebook is a great place to write breakdowns of code and can be stored in a GitHub Repo for others to view and/or collaborate on.

Links & Resources:

1. GitHub Repository: https://github.com/kolbydboyd/jupyter-exploration

Module 02: References

1. Jones, Meghan. "Find the Hidden Objects in These Pictures." Reader's Digest, Reader's Digest, 10 Aug. 2023, www.rd.com/article/find-the-hidden-objects/.
2. Tokoto, Skylar Dayas, et al. "Test: Try to Find the Odd One out in Less than 10 Seconds." Bright Side - Inspiration. Creativity. Wonder., Bright Side, 9 May 2021, brightside.me/articles/test-try-to-find-the-odd-one-out-in-less-than-10-seconds-801866/.
3. Shandilya, Archana. "Word Puzzles Brain Teaser: Identify the Words inside the Boxes through Hints in 30 Secs." Jagranjosh.Com, Jagranjosh.com, 24 June 2022, www.jagranjosh.com/general-knowledge/word-puzzles-brain-teaser-identify-words-inside-boxes-through-hints-clues-1655989947-1.

Tools of the Trade

Module 03

This module introduced machine learning for Computer Vision.

Module 03: Tools of the Trade

Main Learnings:

- Data is the biggest thing in this field. We need data in, we need to test the data, train the machine on the data, then feed it new data to make sure it performs as expected.

Key Takeaways:

- Machine learning is the art of teaching computers to do cool stuff by showing them a bunch of examples, so they can learn to make predictions and decisions on their own, like a digital Sherlock Holmes. (Grok, 2024)

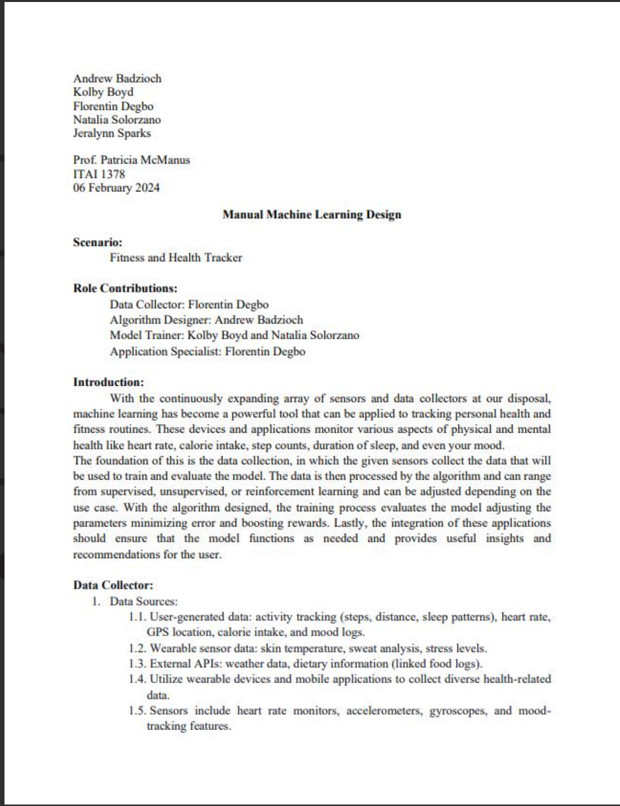
Challenges Faced:

- Jupyter Labs took a while for me to figure out how to set up. I figured it out by doing further study and research on my own with Stack Overflow, Reddit, and good ol’ Googling.

Module 03: Activities & Results

Our group assignment was to document in detail a Fitness & Health Tracker. We individually collected data, visualized algorithms that might be of use, walked through the training process, and applied this knowledge to this real world application.

Fitbit: example of a fitness wearable



Andrew Badzioch
Kolby Boyd
Florentin Degbo
Natalia Solorzano
Jeralynn Sparks

Prof. Patricia McManus
ITAI 1378
06 February 2024

Manual Machine Learning Design

Scenario:
Fitness and Health Tracker

Role Contributions:
Data Collector: Florentin Degbo
Algorithm Designer: Andrew Badzioch
Model Trainer: Kolby Boyd and Natalia Solorzano
Application Specialist: Florentin Degbo

Introduction:
With the continuously expanding array of sensors and data collectors at our disposal, machine learning has become a powerful tool that can be applied to tracking personal health and fitness routines. These devices and applications monitor various aspects of physical and mental health like heart rate, calorie intake, step counts, duration of sleep, and even your mood.
The foundation of this is the data collection, in which the given sensors collect the data that will be used to train and evaluate the model. The data is then processed by the algorithm and can range from supervised, unsupervised, or reinforcement learning and can be adjusted depending on the use case. With the algorithm designed, the training process evaluates the model adjusting the parameters minimizing error and boosting rewards. Lastly, the integration of these applications should ensure that the model functions as needed and provides useful insights and recommendations for the user.

Data Collector:
1. Data Sources:
   1.1. User-generated data: activity tracking (steps, distance, sleep patterns), heart rate, GPS location, calorie intake, and mood logs.
   1.2. Wearable sensor data: skin temperature, sweat analysis, stress levels.
   1.3. External APIs: weather data, dietary information (linked food logs).
   1.4. Utilize wearable devices and mobile applications to collect diverse health-related data.
   1.5. Sensors include heart rate monitors, accelerometers, gyroscopes, and mood-tracking features.

Module 03: References

1. https://www.youtube.com/watch?v=yjjE-MJD5TI&ab_channel=CornellUniversityCenterforAdvancedComputing
2. **Linkedin**: harnessing-power-machine-learning-python-fitness-austin-stewart
   https://scikit-learn.org

Machine Learning Basics

Module 04

This module introduced features and labels for data in Machine Learning. Supervised vs. Unsupervised Learning concepts and applications of each, as well as the ML Lifecycle.

Module 04: Machine Learning Basics

Main Learnings:

- The most common types of data are Tabular (in tables), images, and text.
- Labels are the output that we want to predict
- Features are the input variables/attributes that are used to predict the output

Key Takeaways:

- Supervised Learning: Data is provided with labels and the model learns by looking at the examples
- Unsupervised Learning: Data is provided without labels and the model finds patterns in the data

Challenges Faced:

- Machine Learning requires huge amounts of data, and the data requires lots of preparation.
- ML requires a lot of processing power (GPUs)
- ML uses specialized algorithms, frameworks and libraries

Module 04: Activities & Results

Learning about image classification with k-Nearest Neighbors (k-NN) highlighted the simplicity of algorithms like these and their effectiveness, particularly in classifying handwritten digits using the MNIST dataset.

While k-NN is easy to implement and interpret, some challenges include:

1. Selecting the appropriate value for k
2. Managing computational costs as the dataset size increases

Strategies to mitigate these challenges:

1. Feature selection
2. Cross-validation
3. Ensemble modeling

OmniTech
Kolby Boyd
ITAI 1378
L04

Image Classification with k-Nearest Neighbors

Reflection on Learning:

1. I understand that the k-Nearest Neighbors (k-NN) algorithm is a simple yet effective method for classification tasks. The k-NN algorithm classifies a data point based on the majority class of its k nearest neighbors in the feature space (the n-dimensional points where the variables live).
2. Applying k-NN to the MNIST dataset involved measuring the distances between features that represent images of handwritten numerical digits. The algorithm then classified each image based on the labels of its nearest neighbors.
3. Preparing the MNIST dataset involved loading it, visualizing training samples, and splitting the data into training and testing sets. Training the k-NN classifier was straightforward using sci-kit-learn's implementation. The evaluation involved making predictions on the testing set and assessing its accuracy.
4. One challenge I encountered was choosing an appropriate value for k. I experimented with different values and observed how it affected the model's performance. Insights gained include the importance of feature scaling and the computational cost of making predictions with k-NN.

Responses to Lab Questions:

1. The k-NN algorithm is easy to understand and implement. This makes it a great choice for beginners. k-NN doesn't have a training phase so it can be used for both supervised and unsupervised learning tasks without the need for model training. Predictions made by the k-NN algo can be easily interpreted, because they are based directly on the data. Some cons of the k-NN algo are that as the dataset increases, the computational cost grows significantly, especially since the algorithm stores the entire dataset within its memory. This also increases the prediction time. k-NN is often referred to as "lazy," because it doesn't build a model during the training phase. Instead, it memorizes the training data and makes predictions based on the similarity of new instances to existing instances in the training data.
2. The Implications of this include:
   a. No upfront computational cost due to no training phase.
   b. Best for smaller size datasets for prediction time costs.
   c. Storage must be sufficient depending on the size of the dataset.

Advanced Questions:

1. Using feature selection algorithms or feature importance ranking may reduce the computation time in k-NN for larger datasets. Working with a representative sample of the dataset instead of using the entire dataset may lead to faster computation. This can result in a sample of results to be tested on a more robust model and can be done repeatedly on small sections of the dataset. Using parallel processing techniques in order to distribute the computation across multiple processors may reduce the overall computation time, as well.
2. In k-NN, overfitting can occur when the value of k is too small, leading to a noise sensitive model. On the other hand, underfitting can occur when the value of k is too large, resulting in a model that is too simple and fails to capture the patterns in the data.
3. Some strategies to address the balancing act between underfitting and overfitting include k-fold cross validation to find the optimal value of k that balances bias and variance. Also, experimenting with the combination of multiple k-NN models with different values of k or using different distance metrics to create an "ensemble" model can help to reduce the risk of over- or underfitting.

Module 04: References

1. Pedregosa et al. "Scikit-learn: Machine Learning in Python." Journal of Machine Learning Research 12 (2011): 2825-2830.
2. CS231n Convolutional Neural Networks for Visual Recognition. (n.d.). Cs231n.github.io. Retrieved May 6, 2024, from https://cs231n.github.io/classification/#:~:text=k%20%2D%20Nearest%20Neighbor%20Classifier
3. Python, H. to tune hyperparameters with, & says, scikit-learn- P. (2016, August 8). k-NN classifier for image classification. PyImageSearch. https://pyimagesearch.com/2016/08/08/k-nn-classifier-for-image-classification/

Intro to Neural Networks with TensorFlow

Module 05

TensorFlow

This module introduced us to Tensorflow Playground, an online interactive tool for experimenting with neural networks. It allows us to understand the behavior of neural networks by visualizing and adjusting parameters like the network architecture, activation functions, and dataset characteristics.

Module 05: Intro to Neural Networks

Main Learnings:

- Activation functions determine node outputs, and learning algorithms adjust connections to minimize errors.
- Neural networks (NN) structure and function mimic the human brain

Key Takeaways:

- ReLU activation led to faster convergence compared to sigmoid
- Data noise negatively affects generalization, which means we need robust neural networks.

Challenges Faced:

- Finding the right balance between model complexity and generalization
- Adjusting learning rates to balance the speed of convergence and stability
- Handling data noise to maintain the model’s performance

Module 05: Activities & Results

Neural Networks are computational models inspired by the structure and function of the human brain.

This lab allowed me to visualize the neural network and see what would happen internally if I changed certain parameters.

The differences between activation functions, number neurons and learning rates had major effects on the outputs, as you can see on the next slide.

OmniTech
Kolby Boyd
ITAI 1378
L05

TensorFlow Playground

Introduction to Neural Networks

Neural networks (NN) are computational models inspired by the structure and function of the human brain. They are made up of interconnected nodes organized in layers, including input, hidden, and output layers. Activation functions will determine the output of each node, and learning algorithms adjust the connections between the nodes to minimize errors during training.

Activation Functions

Experimenting with different activation functions revealed notable differences in the network's performance. The ReLU activation function resulted in faster convergence compared to the sigmoid function.

Hidden Layer Neurons

Increasing the number of neurons in the hidden layer initially improved the network's performance, but performance decreased beyond a certain point. Adding more hidden layers also enhanced performance, but the benefits plateaued after a certain amount, indicating the importance of finding the right balance between model complexity and generalization.

Learning Rate

Adjusting the learning rate demonstrated its impact on convergence speed and accuracy. Higher learning rates accelerated convergence but risked overshooting optimal solutions, leading to instability. Conversely, lower learning rates resulted in slower convergence but offered better stability and improved accuracy over time.

Data Noise

Introducing noise in the data negatively affected the network's ability to generalize. As the level of noise increased, the network's performance on unseen data deteriorated. This underscores the importance of robustness in NN models and the need to account for noisy input data during training.

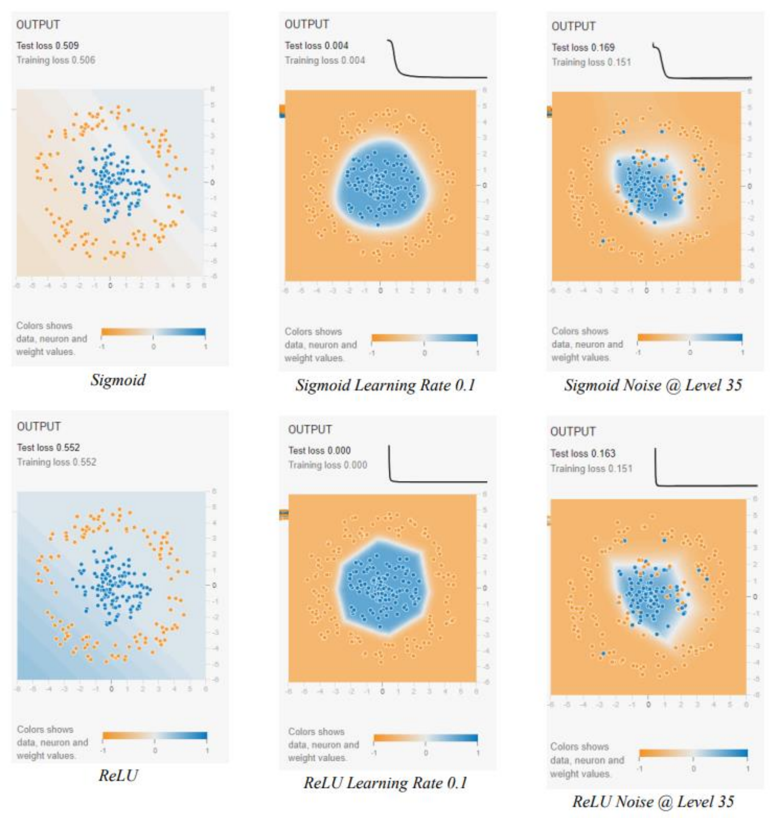
Dataset Explorations

Exploring different datasets highlighted how the characteristics of the dataset will influence network performance. Datasets with clear patterns and less variability yielded higher accuracy, while datasets with complex or noisy patterns posed greater challenges for the network. This emphasizes the importance of dataset selection and preprocessing in neural network applications.

Conclusion:

In summary, this assignment provided valuable insights into the behavior of neural networks through hands-on experimentation. Understanding the impact of activation functions, hidden layer configurations, learning rates, data noise, and dataset characteristics is crucial for designing an effective NN model. There is an immense amount of balancing required to create an effective model. Practical implications include the need for careful parameter tuning and robust testing to ensure optimal performance in the real world.

Module 05: Graphs, Tables, & Images



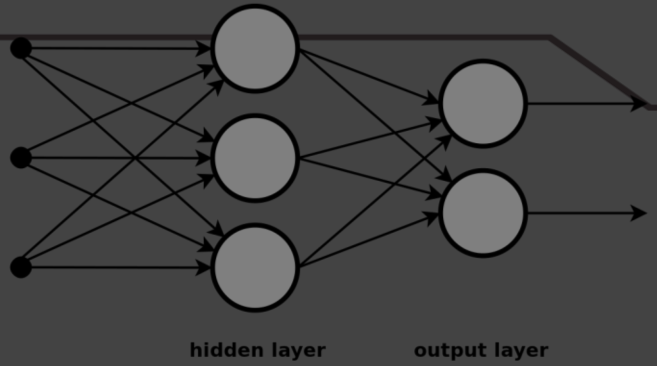
This is the outputs when playing with TensorFlow Playground. I tested many different parameters and learning rates to get these different outputs and visualization of the data.

Module 05: References

1. Carter, D. S. and S. (n.d.). Tensorflow — Neural Network Playground. Playground.tensorflow.org. https://playground.tensorflow.org/#activation=tanh&batchSize=10&dataset=circle&regDataset=reg-plane&learningRate=0.03&regularizationRate=0&noise=0&networkShape=4
2. SHARMA, S. (2017, September 6). Activation Functions in Neural Networks. Medium; Towards Data Science. https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6

Convolutional Neural Networks

Module 06



This first module introduced

Module 06: Convolutional Neural Networks

Main Learnings:

- We can view images mathematically using coordinates
- Image preprocessing techniques:
  - Resizing
  - Cropping
  - Noise Reduction
  - Data Augmentation
- Images can be visualized using a matrix of numbers, with grayscale images having one channel and color images, 3 (RGB).

Key Takeaways:

- Convolutional layers use kernels (sliding filter) to detect patterns in images and produce feature maps
- Parameters:
  - Filter size
  - Stride
  - Padding
  - Activation Functions

Module 06: Activities & Results

Our team drafted a personalized proposal for the implementation of Computer Vision technology to an automobile parts manufacturer. The proposal aimed to revolutionize efficiency, precision, and innovation in Nippo Denso's operations by leveraging Computer Vision for quality control, automated inspection processes, predictive maintenance, and data-driven insights.

Andrew Badzioch
Kolby Boyd
Florentin Degbo
Natalia Solorzano
Jeralynn Sparks

Optic Minds
ITAI 1378
Prof. Patricia McManus
P06

Proposal for Nippo Denso Co. Ltd.
Manufacturer of Automotive Components

Introduction:

We live in a dynamic world where seizing every moment is crucial and contributes to our advancement. We are opening the door to you with another breakthrough solution, Computer Vision. We invite you to explore a groundbreaking solution that promises to redefine efficiency, precision, and innovation in your operations: Computer Vision.

We at OmniTech Enterprises understand the tough challenges you face daily. You are committed to running a world-renowned manufacturing company, producing automotive components. Ensuring your products are manufactured consistently on overly complex production lines takes work. How do you do it?

Definition:

We are proposing Computer Vision. This is how it can revolutionize your operations:

Computer Vision is a technological advancement that uses computers, sensors, and cameras to perceive, interpret, and understand visual information. The capabilities of these systems mirror the human eye but with the highest level of precision and efficiency. For Nippo Denso Co. Ltd., this will mean a massive leap in quality control, inspection processes, and maintenance schedules and will provide new insights into your manufacturing processes.

Benefits:

Consider the advantages awaiting you:

1. **Quality Control:** Identify deviations in the production line as they occur, guaranteeing that only perfect parts reach the assembly line. Computer Vision will limit the number of defects and the need to make corrections.
2. **Automated Inspection Processes:** By automating visual tasks, you can remove the human component of error inspection and speed up production cycles. This will ensure that your Quality Assurance Team can spend their time on more strategic and impactful work.
3. **Predictive Maintenance:** By using equipment monitoring, Denso can predict maintenance needs and prevent costly breakdowns on the assembly line. The stoppage time and maintenance costs will be reduced after implementing predictive maintenance.
4. **Data-Driven Insights:** By employing computer vision technology, the production process can be optimized and streamlined for increased efficiency and cost reduction.

Personal Touch:

Why do we believe Computer Vision is the right move for Nippo Denso Co. Ltd.?

We acknowledge that your successes hinge on innovation, creativity, and flexibility. By embracing Computer Vision, you are continuing to invest in your company's future, not only this modern technology. You are preparing for unprecedented efficiency, trust, and competitiveness in a constantly new market environment.

Conclusion:

Computer Vision is not just a solution, it is a catalyst for transformation. It empowers you to transcend the limitations of traditional manufacturing methodologies and embrace a future where innovation and efficiency converge seamlessly. By welcoming this cutting- edge technology, Nippo Denso Co. Ltd. stands to revolutionize its operations, achieve new levels of precision, and cement its position as an industry leader in

Module 06: References

1. Kelwig, D., & Writer, C. (2023, November 14). 10 sales pitch presentation examples and templates. Zendesk. https://www.zendesk.com/blog/sales-pitch-examples/
2. What Is a Convolutional Neural Network? | 3 things you need to know. (n.d.). Www.mathworks.com. https://www.mathworks.com/discovery/convolutional-neural-network.html#:~:text=A%20convolutional%20neural%20network%20(CNN,-series%2C%20and%20signal%20data.
3. Saha, S. (2018, December 15). A Comprehensive Guide to Convolutional Neural Networks—the ELI5 way. Towards Data Science. https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

Basic Architecture & Transfer Learning

Module 07

This module introduced CNN core architectures & transfer learning. We experimented with k-NN classifiers and explored handling datasets.

Module 07: Basic Architectures...

Main Learnings:

- Evaluating model performance is paramount. This includes techniques such as early stopping, and analyzing the metrics at each step.

Key Takeaways:

- k-NN classifiers must have an optimal k value for improved generalization
- GridSearchCV does this automatically, aiming to minimize the effects of overfitting and underfitting.

Challenges Faced:

- Long compute time for training the k-NN classifiers
  - Tried parallel processing for faster computation
- Addressed low accuracy scores by exploring:
  - Hyperparameter tuning
  - Further normalizing data
  - Extensive preprocessing

Module 07: Activities & Results

This Lab taught me hands on experience with training machine learning models such as k-Nearest Neighbors (k-NN) classifiers by exploring different values of k and using cross-validation for model evaluation.

I also learned techniques for handling and preprocessing datasets, such as flattening labels and performing data normalization to improve model performance.

Kolby Boyd

ITAI 1378

Prof. McManus

L07

Object Detection Lab

Pre Lab Setup:

I downloaded the notebook from the course page on Canvas and chose to run it in Google Colab for convenience. Then, I installed all of the necessary libraries and modules. These include: TensorFlow, NumPy, Scikit-Learn, and MatPlotLib.

Luckily, I did not need to install additional libraries or modules as they were all included in the Notebook.

Model Implementation:

I loaded the CIFAR-10 dataset using TensorFlow's Keras API. After loading, I encountered an error in the prewritten code. I needed to load and unpack the CIFAR-10 dataset using the 'load_data()' function from the 'tf.keras.datasets.cifar10' module. This should return four values: 'x_train', 'y_train', 'x_test', and 'y_test'. My updated code is: '(x_train, y_train), (x_test, y_test) = tf.keras.datasets.cifar10.load_data()'.

After this step, we needed to flatten the labels. This means reshaping the label arrays from a 3D shape into a 2D shape. In the CIFAR-10 dataset, the labels need to be flattened for convenience and compatibility with the machine learning algorithms and loss functions. A 2D array is much easier to work with than a 3D array and Scikit-learn's API is designed to work seamlessly with 2D arrays for labels.

Next, preprocessing of the data. Scaling the data to have a mean of 0 and a variance of 1 (we do this with the 'StandardScaler' in Scikit-learn), is a common preprocessing step in machine learning. This process is also known as z-score normalization or standardization. Standardization centers the data around zero, which can help to remove any bias and ensure that each feature has a similar scale. This process also increases the model performance for algorithms that are sensitive to the scale of the features (such as k-NN or k-nearest neighbors).

Finally, using the 'display_images' function, we learned to display images along with their labels, which is helpful for visual verification of the data. This function is helpful for visually inspecting the images in the dataset, which allows us to verify that the images and labels are loading correctly and processing true, before proceeding with further analysis.

Module 07: References

1. Srivastava, T. (2018, March 25). A Complete Guide to K-Nearest Neighbors (Updated 2023). Analytics Vidhya. https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/#:~:text=KNN%20classifier%20operates%20by%20finding
2. IBM. (2023). What is the k-nearest neighbors algorithm? | IBM. Www.ibm.com. https://www.ibm.com/topics/knn

Advanced Architectures & Object Detection

Module 08

This module taught me real world application of the previous concepts into a model designed to detect between a muffin or a chihuahua.

Module 08: Advanced Architectures...

Main Learnings:

- We have gone through a lot of stuff up to this point! Our cheat sheet ended up being more than one page.
- This time, I used Google Collab GPU 4 to run the model.

Key Takeaways:

- I achieved a 75.84% accuracy by the end of my testing. I'm sure it could be better.

Challenges Faced:

- Fitting everything that we have learned onto one page was most difficult.
- Training a model to identify between a chihuahua or a muffin took a while to get the model tuned but in the end I was successful.

Module 08: Activities & Results

This was our cheat sheet for this class. We created it to have a brief overview of the various concepts we have been learning up to this point.

Andrew Badzioch
Kolby Boyd
Florentin Degbo
Natalia Solorzano

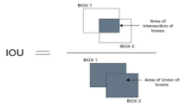
Optic Minds
ITAI 1378
Prof. Patricia McManus
A08

A08 ITAI 1378 CV _Cheat sheet creation

Object Detection Cheat Sheet

Key Concepts:

- Bounding Box: A rectangle drawn around an object in an image.
- Annotations: Labels that identify the objects within bounding boxes.
- Confidence Score: A measure of how confident the model is that an object is present in a bounding box.
- Intersection over Union (IoU): A metric used to evaluate the accuracy of object detection models by measuring the overlap between predicted and ground truth bounding boxes.



Common Algorithms:

- R-CNN: Regions with Convolutional Neural Networks
- Fast R-CNN: Improves on R-CNN by using a single convolutional network to process the entire image and sharing computation between region proposals.
- Faster R-CNN: Improves on Fast R-CNN by using a Region Proposal Network (RPN) to generate region proposals.
- SSD: Single Shot Multibox Detector, a faster alternative to R-CNN that performs object detection in a single pass.
- YOLO: You Only Look Once, another fast alternative to R-CNN that treats object detection as a regression problem.
- Stereo Matching: Uses stereo images (images taken from two viewpoints) to estimate the depth of objects in a scene.
- Monocular Depth Estimation: Estimates depth from a single image using techniques like learning-based methods or geometry-based methods.

Module 08: References

1. (PDF) comparison of Yolo v3, faster R-CNN, and SSD for real-time ... (n.d.). https://www.researchgate.net/publication/353590069_Comparison_of_YOLO_v3_Faster_R-CNN_and_SSD_for_Real-Time_Pill_Identification
2. Base, T. (2024, January 19). AWS Certified Machine Learning Cheat sheet-built in algorithms 2/5. Medium. https://medium.com/@tantabase/aws-certified-machine-learning-cheat-sheet-built-in-algorithms-2-8-cfb70b60603e
3. Convolutional Neural Networks cheatsheet star. CS 230 - Convolutional Neural Networks Cheatsheet. (n.d.). https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-convolutional-neural-networks
4. Sojasingarayar, A. (2024, February 23). Faster R-CNN vs Yolo vs SSD-object detection algorithms. Medium. https://medium.com/ibm-data-ai/faster-r-cnn-vs-yolo-vs-ssd-object-detection-algorithms-18badb0e02dc

Advanced Topics in Object Detection

Module 09

This module introduced advanced techniques in the field of object detection architecture. We learned about transfer learning in this module as well.

Module 09: Advanced Topics...

Main Learnings:

- Transfer Learning leverages a pre-trained model and adapts it to our specific task
- Object detection localizes and classifies multiple object within an image vs. Image classification which assigns a label to an entire image

Key Takeaways:

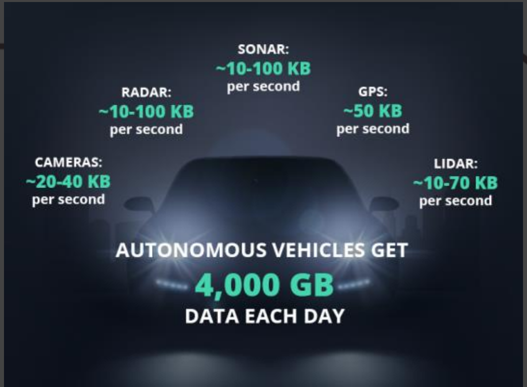
- ChatGPT APIs can be used for our ouwn businesses
- Transfer learning works best when the source & target tasks are similar (image recognition)
- Still the need to fine-tune data

Module 09: References

1. OpenAI API. (n.d.). Platform.openai.com. https://platform.openai.com/
2. Donges, N. (2022, August 25). What is transfer learning? Exploring the popular deep learning approach. Built In. https://builtin.com/data-science/transfer-learning
3. Object Detection: Models, Architectures & Tutorial [2022]. (n.d.). Www.v7labs.com. https://www.v7labs.com/blog/object-detection-guide
4. Kundu, R. (2023, January 17). YOLO: Real-Time Object Detection Explained. Www.v7labs.com. https://www.v7labs.com/blog/yolo-object-detection

Computer Vision for Video

Module 10



This module introduced us to Amazon Rekognition and the understanding of video & video processing in computer vision.

Module 10: Computer Vision for Video

Main Learnings:

- Video data is a sequence of images with visual information played in rapid succession
- Involves not only analyzing the individual frames, but also the temporal relationships between the images.
- Optical flow algorithms estimate the motion of objects in consecutive frames by analyzing pixel intensity changes (enables motion detection and video stabilization)

Key Takeaways:

- Feature extraction still works to identify patterns and structures like edges
- Video data processing requires techniques accounting for temporal dynamics and changes over time
  - Optical Flow
  - Motion Analysis
- Poses challenges such as compute complexity, large amounts of data, and ensuring robustness

Module 10: Activities & Results

This paper is a reflection on a recent workshop discussing Amazon Rekognition, a deep learning image and video analysis service. We highlight its capabilities such as object and facial recognition, emotional detection, and text extraction.

Potential Benefits in sectors such as:

- Law enforcement
- Business

Andrew Badzioch
Kolby Boyd
Florentin Degbo
Natalia Solorzano

Prof. Patricia McManus
ITAI 1378
08 April 2024

A10 Amazon Rekognition guest presentation reflection

At a recent workshop, the speaker explored the practical benefits of Amazon Rekognition, a technology that can recognize faces, objects, and text within images and videos.

Amazon Rekognition is a deep learning image and video analysis service provided by Amazon Web Services (AWS). Some technical aspects include:

1. Amazon Rekognition can identify objects, scenes, and videos which enables applications to automatically categorize visual content.
2. Rekognition is able to identify faces, estimate age range, detect emotions, and recognize facial attributes such as gender or facial hair with its facial detection and recognition capabilities. Rekognition can even identify celebrities!
3. Users can train the model using custom labels to recognize specific objects or scenes relevant to their own applications.
4. The service can detect and extract text from video and images making it useful for applications involving optical character recognition.

This powerful tool offers a profound insight into multimedia content on a large scale, proving invaluable in various real-world applications. From aiding law enforcement agencies in identifying suspects to assisting businesses in gauging customer sentiments through facial analysis, Amazon Rekognition has demonstrated versatility across different sectors. However,

Module 10: References

1. "What Is Amazon Rekognition? - Amazon Rekognition." Docs.aws.amazon.com, docs.aws.amazon.com/rekognition/latest/dg/what-is.html.
2. Kaur Gill, Dr. Jagreet. "Amazon Rekognition Benefits and Its Use Cases." Www.xenonstack.com, 29 Oct. 2022, www.xenonstack.com/blog/amazon-rekognition-benefits.
3. Anthony. (2023, May 7). Amazon Rekognition: How to use Amazon's facial recognition technology for Advanced Analytics. Signalytics. https://signalytics.ai/amazon-rekognition/
4. The Facts on Facial Recognition with Artificial Intelligence. (2019). Amazon Web Services, Inc. https://aws.amazon.com/rekognition/the-facts-on-facial-recognition-with-artificial-intelligence/
5. A governance framework for algorithmic accountability and transparency. (2019). https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf

Generative AI for Computer Vision

Module 11

This first module introduced

Module 11: Gen. AI for Computer Vision

Main Learnings:

- GANs are two neural networks, a generator and a discriminator which compete against each other to generate realistic samples and distinguish between real and generated samples simultaneously.
- Autoencoders learn to encode input data into a lower (pixelated) resolution and decode it back again to reconstruct to original data, which can be used to generate new samples.

Key Takeaways:

- Generative models can create new data samples
- Creative applications include:
  - Image generation
  - Style transfer
  - Artistic manipulation
- Ethical Considerations:
  - Data privacy
  - Authenticity of works
  - Potential for misuse

Module 11: References

1. How Generative AI in Computer Vision Drives Productivity. (n.d.). Www.xenonstack.com. Retrieved May 6, 2024, from https://www.xenonstack.com/blog/generative-ai-in-computer-vision#:~:text=Generative%20AI%20is%20revolutionizing%20computer
2. chooch. (2023, June 29). 4 Ways Generative AI Is Improving Computer Vision | Chooch. Vision AI | Generative AI for Computer Vision | Chooch. https://www.chooch.com/blog/4-ways-generative-ai-is-improving-computer-vision/
3. Generative AI in computer vision: Use cases and benefits. (n.d.). Www.softwebsolutions.com. Retrieved May 6, 2024, from https://www.softwebsolutions.com/resources/generative-ai-in-computer-vision.html

Autonomous Systems for Computer Vision

Module 13

In this module, we learned about AI agents that are with us everyday. These systems learn from us and understand us well enough to assist our day-to-day activities.

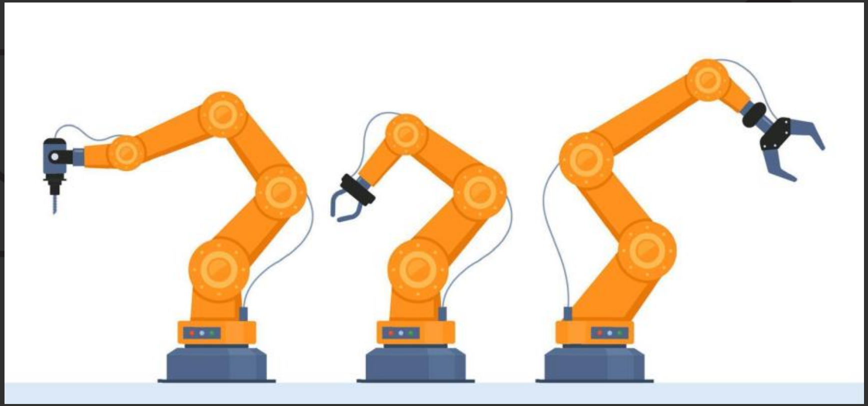
Module 13: Autonomous Systems...

Main Learnings:

- AI Agents: invisible helpers (computer programs) that can make decisions, learn, and understand our needs.
- Examples include:
  - Virtual assistants
  - Customer service chatbots
  - Factory robots
  - Self driving cars
- Intelligent agents can be:
  - Reactive, responding to stimuli, such as a Roomba robot
  - Deliberative, planning actions based on goals such as an AI playing GO.
  - Hybrid, combining reactive and deliberative elements such as a self driving car with collision avoidance and route planning

Key Takeaways:

- **Computer vision is a key technology that allows these agents to perceive their surroundings**



Module 13: References

1. Schlicht, M. (n.d.). The Complete Beginners Guide To Autonomous Agents. Matt Schlicht's AI Newsletter. https://www.mattprd.com/p/the-complete-beginners-guide-to-autonomous-agents
2. Johnston, J. (2002). A Future For Autonomous Agents: Machinic Merkwelten and Artificial Evolution. Configurations, 10(3), 473–516. https://muse.jhu.edu/article/51658
3. What are AI Agents?- Agents in Artificial Intelligence Explained - AWS. (n.d.). Amazon Web Services, Inc. Retrieved May 6, 2024, from https://aws.amazon.com/what-is/ai-agents/

Conclusion & Future Objectives

Learning about machine learning and its application to computer vision enabled me to understand how computers can analyze and interpret visual data, leading to advancements in fields like facial recognition, autonomous vehicles, and medical imaging.

It was difficult at times but overall the challenges were exciting as I learned to work through the problems.

I enjoyed working alone and with my team members as well as all of the extracurricular opportunities for networking.

My next steps are to finish summer fall and spring semesters to graduate with an AAS Artificial Intelligence.

At this stage, robotics software or systems engineer looks promising as a career.

I enjoy working with my hands but I also love the intricacies ad complexity of Big Data.

Thank You!

Let’s bring on Summer & Fall 2024!