CS GY 6643 - Computer Vision, Fall 2024 **Guido Gerig**

Homework 3:

Out: Thu April 11-2024
Due: Thu April 18-2024
This is a pencil/paper homework that we use for reviewing materials taught in our course. The workload is timed to be about 1-2hrs.

Late submissions: Late submissions result in 10% deduction for each day. The assignment will no longer be accepted 3 days after the deadline.

Office hours:

		Mon	Tue	Wed	Thur	Fri
Guido Gerig	gerig@nyu.edu				2-3pm (ZOOM)	
Pragnavi Ravuluri Sai	pr2370@nyu.edu					8am – 10am EST
Sai Rajeev Koppuravuri	rk4305@nyu.edu			12.00-13.00 EST		

Please remember that we also use campuswire for communication on homeworks.

Please submit a **pdf document** where you combine the questions with your answers. You can use pencil&paper for handwritten answers or some other tools for graphing of plots or lists, then scan all pages with an app such as CamScanner or similar, and finally combine into a single pdf document. Please make sure that your answers are well readable for our grading (contrast, hand-writing, etc.).

This is a project to be solved by **each student individually**. Solutions and reports that may indicate copying materials from other students or from web resources are considered plagiarism and subject to violation of the honor code.

---

1. Deep Learning (

Questions related to guest lecture by Axel Elaldi, discussions during his lecture, and his sets of slides that are uploaded to Brightspace.

1.1 Briefly explain the difference between classical shallow learning and deep learning.

Please be brief in your explanations. Short, itemized sentences are good enough.

Classical Learning

………….
………….
…………

Deep Learning

…………..
…………..
………

1.2. Discuss the Pros and Cons of a CNN over an MLP architecture.

Also, describe what happens to the input pixel of the MLP and a CNN when a translation is applied to the image.

Please be brief in your explanations. Short, itemized sentences are good enough.

MLP:
………
………
………

Under Translation, …….

CNN:
………
………
………

When Translation of the image is applied, ………

2. Perspective Projection

2.1. Starting from the basic perspective equations for a pinhole camera, $x = f\frac{X}{Z}$ and $y = f\frac{Y}{Z}$, where (x,y) are coordinates in the camera and (X,Y,Z) the coordinates of a 3D point, briefly explain the fundamental problem of 3D reconstruction of a 3D point in a scene given a 2D point in your camera

The perspective projection equation shows that 3D points in the world get projected to single **p=(x,y)** points on the sensor. The primary goal of computer vision is the inverse, i.e. to use image points to reconstruct the 3D point location **P=(X,Y,Z)**. With only one camera, we would use the ray through (x,y) and the optical center into the world, but don't know where the world point came from, since all points on this ray (green dots in the image below) would be projected to the same **p=(x,y)** location. With a stereo setup and successful detection of corresponding points, we can find a solution for **P**.

2.2. Given the following scenario: Given is an object in the scene that you know is 3m high (e.g. a truck of known dimensions along the vertical axis), its camera image showing a vertical dimension of 20mm, and the focal length of 50mm. Calculate the distance of the object from the camera if you think you can.

2.3 Given your answer to 2.1 before and now to 2.2, briefly discuss why there is (is not?) a fundamental problem in scenario 2.1 but then a solution (or not) in 2.2.