Neural Networks

A pan-tilt-zoom (PTZ) camera is a robotic camera capable of panning horizontally (from left to right), tilting vertically (up and down), and zooming (for magnification). PTZ cameras are often positioned at guard posts where active employees may manage them using a remote camera controller. Their primary function is to monitor expansive open regions that need views in the range of 180 or 360 degrees. See the Wiki page and this YouTube video to understand more about PTZ cameras. PTZ cameras can be installed in indoor environments, out-door environments, and on various vehicles such as drones.

a) Out-door PTZ camera

b) PTZ camera on a drone

Figure 1. PTZ cameras installation on different environments.

PTZ cameras must be controlled manually or by autonomous algorithms or combined. In this question, you will propose a Deep Learning based approach for autonomous control of PTZ camera where your system will also allow classification of the objects of interest. Indeed, your camera will orient itself to a default position where no interesting object is in the scene. Once an object of interest is detected the PTZ camera will change its orientation to focus on the object and classify it. You can propose a single end-to-end neural network that does both camera control and classification simultaneously or these 2 systems can be different architectures. Your proposal must contain at least one deep learning architecture. You can use or inspire from the neural network architectures we covered in the course, or you can suggest your own architecture. In both cases, necessary details must be provided. You do not have training data and your GPU resources are limited. So come up with some clever idea that will work with this limitation.

Give:

- Your proposals system diagram,
- Details on employed neural network architectures,
- How you will obtain training data if training data is necessary,
- How will you train it? Using CPU or GPU? Optimizers? Parameters? Etc.
- How will you control the PTZ camera using your system?
- How will you classify the focused scene?
- Give your reasons, do not just say I am using xxx architecture, explain why you specifically preferred that one. What are the advantages and disadvantages of the architecture you choose?