3. Computer Vision Task

Hand-pose estimation: Suppose you are given a dataset of 150k images of random (but exhaustive) hand poses along with the finger joint positions (3D coordinates against 3 joints in each of the 5 fingers) for all 150k images. You have to train a ResNET to be able to predict the 3D joint locations given a hand pose (resembling the image format of the dataset).

1. What kind of machine learning task is this?

Since the full research paper was not open source and available anywhere for free (at least not in my knowledge), so the answer is based on what is available on the YouTube video and the abstract of the research paper visible on the link.
This is semi-supervised machine learning because first the thermal images from different angles are joined or stitch by a neural network and then 20 joints in 3D space is also found by the neural network. The transformation of 2D thermal images into 3D perception can be considered the unsupervised. If the transformation is unsupervised which it seems to be so this is an example of semi-supervised learning.

2. How will you execute the paper? Attempt to write the execution pseudo-code.

IMPORT ResNET
IMPORT FC_layers
SET data pipeline
SET neural network
SET hyperparameters (learning_rate, number of FC_layers)
DEFINE metrics
REPEAT training
UNTIL specific metrics values are achieved
or
UNTIL neural net is converged

3. What possible hurdles do you see in this project?

It seems very easy to follow project but perhaps the biggest challenge is the actual training time because the dataset is large and training a ResNET 34 on such large network will take a lot of time. Not to mention, fine-tuning, pruning and optimization of the neural net also take more time. Other than that I have some technical critical reviews about the approach used in the paper that is instead of using FC layers at the end they could have used ***GCN*** layers which normally outperforms these traditional FC layers meaning resulting in higher performance and quality.

4. How will you document your efforts and report progress during training your implemented model?

– Data split into training, validation and testing.
– A mosaic of each split in the document.
– Training and testing metrics progress with each epoch or iteration in the form or graphs or in the form of some sort of data-loggers (like wandb).
– The end results on training, validation and testing data comparing the results with the ground truth values.
– Further improvements that can be done to improve the results of this network.

In the code, I will map the functionality of the code with the sections of research paper or the implementation of the paper. This helps the programmer to understand the code without looking back and forth into the paper again and again. I usually call it the *shorthand* for coder.