**DEEP LEARNING PROJECT PLAN SUMMARY**

* **Topic**
* **Research objective / context and motivation**

We will try to classify sounds coming from urban areas. We would like to be able to detect abnormal sounds such as gun shots among more common ones. The idea would be to help the city administration/police detect more quickly dangerous events…

* **Methodology**

The general idea is to transform sound data into images through feature engineer technique such as “Mel” spectrogram (i.e. scale base on pitch) or chromagram (i.e. scale base on pitch categories). This manipulation will allows us to apply a convolutional Neural Networks CNNs and use each feature engineer as a channel to end up with a feature maps. RNN is also a possibility.

* **Dataset**
  + Music Analysis Dataset : <https://github.com/mdeff/fma>
  + Million Song Dataset : <http://millionsongdataset.com/>
  + Open speech : <http://www.openslr.org/12/>
  + VoxCeleb : <http://www.robots.ox.ac.uk/~vgg/data/voxceleb/>
  + Urban Sound Classification : <https://datahack.analyticsvidhya.com/contest/practice-problem-urban-sound-classification/>
  + AudioSet : <https://research.google.com/audioset/?fbclid=IwAR3If9WF29_QwarlvzjwylQVYxxTKNhCAcpA0vanD_hhOe0e8XVfVcyFMYs>
* **Data preprocessing**
  + Need to do some literature review in order to tackle that feature engineering. Could we use a model that already extracts some of these features automatically (AutoInt ?)
  + For the preprocessing, there are many different articles on how to transform sound files into direct spectrogram. <https://fairyonice.github.io/implement-the-spectrogram-from-scratch-in-python.html> This one is a good example. Also this one <https://stackoverflow.com/questions/44787437/how-to-convert-a-wav-file-to-a-spectrogram-in-python3>
  + There is a possibility for us to use data augmentation to make sure our dataset is bigger. Not only could it make it easier to train with more example, but depending on how we decide to do this it could also account for distortion in the sound received through the means we have. Make the model more robust.
    - This paper has an interesting idea explaining how we can use methods for data augmentation both on the data directly and on the spectrogram we created after: <https://arxiv.org/ftp/arxiv/papers/1912/1912.05472.pdf>
  + When it comes to very specific examples of how we can do the data augmentation there are a few methods in this paper. If we take a good look at how they do it we can also create multiple different sets of data so that we can figure out what the optimal data augmentation type is or maybe a hybrid. <https://arxiv.org/pdf/1608.04363.pdf>
* **Estimation method**
  + Leverage some of the feature extraction or feature engineering by using pre-trained CNN model such as the VGG-16 model and then train a fully-connected network.
  + Based on the graph we can see in this paper we can see that an architecture we can decide to use is CBAM which could allow us to use attention in the estimation. It is also very recent : <https://arxiv.org/ftp/arxiv/papers/1901/1901.06032.pdf>
  + When it comes to a good architecture for models there is this paper here that gives us an overview of what they use for acoustic data. The type of data is slightly different but the architecture could still help us For their state-of—the-art they use BoAw, DNN+HMM. Additionally, their optimal architecture is two conv layer followed by 1 pooling 3 times and 3 FC layers after. <https://arxiv.org/pdf/1604.07160.pdf>
  + There is also this paper that goes through most of the ones we know and love like ResNet <https://arxiv.org/pdf/1609.09430.pdf> Can help us at least get an idea of how we should look for them.
* **Hyperparameters tuning**
  + Recommendation about the type of hyperparameter we should focus on vs the state-of-art default parameter.
* **Performance measure**
  + Is the accuracy still a good indicator?
* **References**
  + <https://ieeexplore.ieee.org/document/7324337>
* **Questions**

1. What is the best architecture for sound classification? Is the transformation into images something we should consider doing? Are there any other methods known?
2. Do you see any challenges with our project? What would you like to ~~see~~ hear in the final report regarding our project?