Discussion agenda for Sprint 1 – Week 2

* YouTrack
* Stories for Sprint 2 created – I will create some tasks after Sunday and decision on what all we will be working on? (Manju – due 6/8)
* Action Item – all team members
  + Finish all of your Sprint 1 items and move items off to Resolve state (Due 6pm Sunday 6/2)
  + Review items not belonging to you in “To Verify”, add your name as reviewer and move to Done (Due 6pm Sunday 6/2)
* Discuss Project Scope
* Algorithms for Image detection and Image classification

|  |  |  |
| --- | --- | --- |
| **Problem** | **Exist on Kaggle kernels/Medium Blog/GitHub** | **We can explore below architectures** |
| Image Classification | CNN | 1. CNN   2. Hybrid CNN-ELM - Image Classification  3. Mask R-CNN - Image Segmentation |
| Object detection | Faster R-CNN | 1. Faster R-CNN with different Hyper parameter  2. YOLO (You only look once)  3. SSD (Single shot Multi-box detection) |

* Additional stretch scope items that team has committed to
  + Building UI or Visualization (to clarify with Prof)
  + Standardized Framework (Discuss with Prof so we are on the same page of what we mean by this) – capture work and if it can be reused by others.
  + Cloud (check with Prof for value to project)
* Define Data needs
* Data exploration (review datasets visually to understand quality) (Manju – create story and tasks – due 6/2)
* Data for elemental image detection and classification
  + Existing data suffices for non-Mask R-CNN algo.
  + May not be adequate for Mask R-CNN (Check with Prof)
* Data for stretch scope
  + Search for another dataset for Std framework (future sprint)
* Project plan - Manju
  + Will keep below ready for tomorrow (Manju – due 6/2)
  + High level stories per Sprint
  + High level task items per Sprint
  + Dates associated with high level task items
  + Will mimic YouTrack board to above
* PowerPoint –
  + Update PowerPoint format/template and add additional high-level items to cover during Tue class (Manju – update by tonight 6/1)
  + ~~Organization of material (Yinchen to maintain OneDrive copy….will maintain Sprint presentation copies)~~
  + Github repo for all shared material
* Word Write-up
  + Update write-up to include additional items.
  + ~~Organization of material (Z to maintain OneDrive copy….will maintain Sprint presentation copies)~~
  + GitHub repo for all shared material
* Technology
  + Create Github project (share details – Manju 6/1)
  + Required software installation on laptops
    - Identify software and confirm it is installed (All team members target: 6/6/19)
      * Jupyter notebook
      * Python 3.7.2
      * Keras
      * Tensorflow
      * PyTorch
  + Identify and understand use of Argo (target review with Jeff : 6/15/19)
    - Submit request for Argo
    - Including use of Swarm and executing jobs (can start with a test job)
    - Explore GitLab on Argo
  + Identify additional software needed for stretch scope items
  + Building UI or Visualization (to clarify)
    - Explore SSD MobileNet for a live demo.
    - UI framework – JS (TBD)
  + Standardized Framework
    - TBD
* Share Reading ideas/add to references section
  + Exploring Algorithmia.io (Manju)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 5/21 | 5/28 | 6/4 | 6/11 | 6/18 | 6/25 | 7/2 | 7/9 | 7/16 | 7/23 | 7/30 | 8/4 |
| Problem Definition | |  |  |  |  |  |  |  |  |  |  |
|  |  | **Data Exploration** | |  |  |  |  |  |  |  |  |
|  |  | Image result for star**Technology Exploration/ Setup** | | |  |  |  |  |  |  |  |
|  |  |  | | **Core Project Scope - Development** | | | | |  |  |  |
|  |  |  | |  | **Stretch Scope Projects items - Development** | | | | |  |  |
|  |  |  | |  |  |  |  |  |  | **Final review/ Presentation** | |

**Project Timelines:**

**Detailed Project Plan (work-in-progress):**

|  |  |  |
| --- | --- | --- |
| Milestone Goals | High level activities | Date |
| Data Exploration |  | **Due 6/18** |
|  | **Understand provided dataset** |  |
|  | **Determine data preparation required and associated activities** |  |
| Technology Exploration/Setup |  | **Due 6/25** |
|  | **Identify individual laptop software requirements** |  |
|  | **Explore Argo** |  |
|  | **Identify software/algorithms to learn and start learning** |  |
| Core Project Scope |  | **Due 7/16** |
|  | **For algorithms in scope, start development** |  |

Expected achievements for Sprint 2 – Week 1

* Data exploration
* Laptop setup
* Argo exploration
* Understanding of basic algorithms

Discussion agenda for Sprint 2 – Week 1

**Data Exploration:**

Table showing all characters with over 100 images



**Example images of Homer Simpson showing complexity of image classification:**

Image with Homer Simpson and multiple other characters



Image showing Homer Simpson in different angle (face obscured)



Similar sized characters in an image



Homer Simpson with head wear and body obscured



**Other items to add to Sprint 2-Week1:**

Add Z’s graph

Using Template to write data exploration part of report.

Individual laptop setup – status

* Anaconda based install
* Downgrade Python to 3.6
* Create environments
* Install keras
* Install PyTorch (different environment)

Argo cluster exploration – status

Explore TensorBoard (<https://www.tensorflow.org/guide/summaries_and_tensorboard>)

**Sprint 2 - Week 2**

Items to capture in presentations:

What did you accomplish?

What did you try to accomplish?

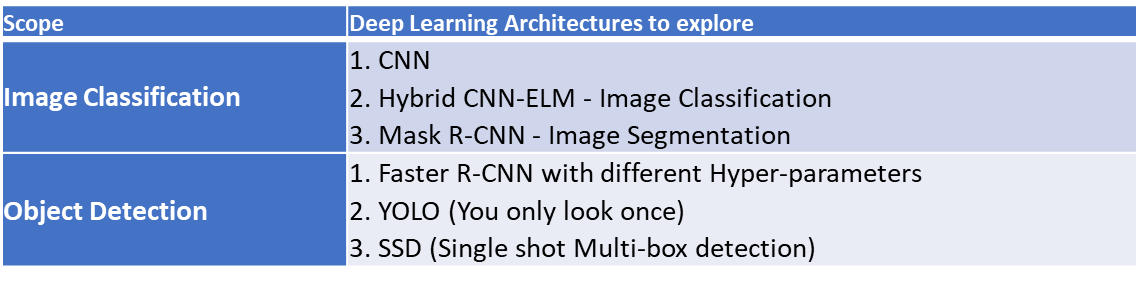
What did you plan for and didn’t accomplish?

What is next?

Data Quality assessment – due Sprint 2 – Week 2

YouTrack screenshots

Expected achievements for Sprint 2 – Week 2

* ~~Data exploration~~
* ~~Laptop setup~~
* Argo exploration
* Understanding of basic algorithms
  + Papers/references
  + 
  + All of us read about these 6 and on Sunday have learnings to share

Sprint 3 – Week 1:

* Make repo private - Manju
* Move all artifacts to GitLab – Manju
* Create high level stories in YouTrack – Manju
* Build out Report – Y & Z
* Data prep:
  + Same data prep applicable for all algorithms?
* Algorithm development:
  + Ravi
    - Faster R-CNN
    - Hybrid – CNN ELM
  + Z & Y
    - CNN
    - Mask R-CNN
  + Manju
    - YOLO
    - SSD
* Data has been uploaded to GitLab
* CNN run with all data took about 13 minutes in GPU node 50 (Test loss: 0.8010630374667288, Test accuracy: 0.8317659352972548)
* Update on Learnings

Sagemaker hyperparameter approaches:

Grid Search

Random Search

Bayesian optimization