Какие признаки мы оставляем для обучения

* Habitat – Среда обитания, важный признак
* **Substrate – крутой признак, указывает на какой почве растет гриб**

**НЕ ОСОБО ВАЖНЫЕ ПРИЗНАКИ, НО ПУСКАЙ БУДУТ ХОТЯ НУЖНО ЕЩЕ ПРОВЕРИТЬ**

* Elevation – высота над уровнем моря
* landcover – растительный покров
* biogeographicalRegion - биогеографический регион

Какие признаки мы удаляем

* eventDate, year, month, day – всяки различные форматы дат снимка, смысла в этом нет
* countryCode – смысла нет, это изучения локации, типо Канада, США и прочее
* scientificName – это научное название, смысла нет
* hasCoordinate – имеются ли координаты, да или нет, смысла нет
* species – смысла нет, так как это просто очень корреляцинно связанный признак с specificEpithet
* **iucnRedListCategory – можно подумать, но пока не вижу применения этого признака**
* latitude, longitude, coorUncert, region, district – указывает на местоположение объекта в соц мире
* observationID – id уникального объекта (гриба), то есть в тесте мы должны будем взять этот id и благодаря нему мы сможем использовать аж несколько фото для определения класса
* filename – название фото
* metaSubstrate – более подробное описание почвы, считаю что не нужно нам. ХОТЯ ПОД ВОПРОСОМ

Признаки для предсказывания

* category\_id – САМЫЙ ВАЖНЫЙ ПРИЗНАК ПО КОТОРЫЙ МЫ И БУДЕМ ОТПРАВЛЯТЬ НА ПРОВЕРКУ В KAGGLE, ВСЕ ПОСЛЕДУЮЩИЕ БУДУТ НУЖНЫ ТУПО ДЛЯ ОБУЧЕНИЯ
* kingdom – важный признак, так как у нас задача ранжирование, показывает царство в которое входит определенный id, разные id могут иметь одинаковое царство
* phylum – важный признак, указывает на тип растения, в которое входит определенный id, разные id могут иметь одинаковое царство
* class – класс растения, в которое входит определенный id, разные id могут иметь одинаковое царство
* order – уже более точный признак, указывает куда более большую принадлежность к определенному грибу

**Тут уже можно задаться вопросом а стоит ли еще сильнее углубляться и усложнять сеть**

* family – семейство, еще более точный признак
* genus – и еще более точный признак
* specificEpithet – еще более определенный признак

Создаем архитектуру модели

У нас есть observationID – это номер уникального объекта, у одного объекта может быть несколько изображение, ну то есть несколько векторов. И это все случайно

Теперь задача стоит, как лучше поступить с этими патчами, сначала я решил попробовать механизм внимания, что мы делаем, допустим 5 фото, мы их пропускаем через сеть CNN, после получаем карты, их мы объединяем, и пропускаем через механизм внимания, это для фото.

А для доп параметров, мы делаем ровно то же самое, только один механизм внимания работает на выявление важных изображение, а вторый механизм внимания будет работать на важность соотношения фото/доп параметров

Я написал архитектуру модели, я взял готовую архитектуру resnet18 продобученную, после, создал свою архитектуру, что было проделано

Как будет выглядить архитектура

Resnet18-> Attention photo-> один CNN-> один входной признак-

У нас изначальный x имеет размерность [num\_images, chanel,width,height] добавляем батч [batch, num\_images, chanel,width,height], pytorch обрабатывает только размерность 4, без num\_images, получается нам это нужно с чем то объединить, с батчами, глупо, затеряется, а вот