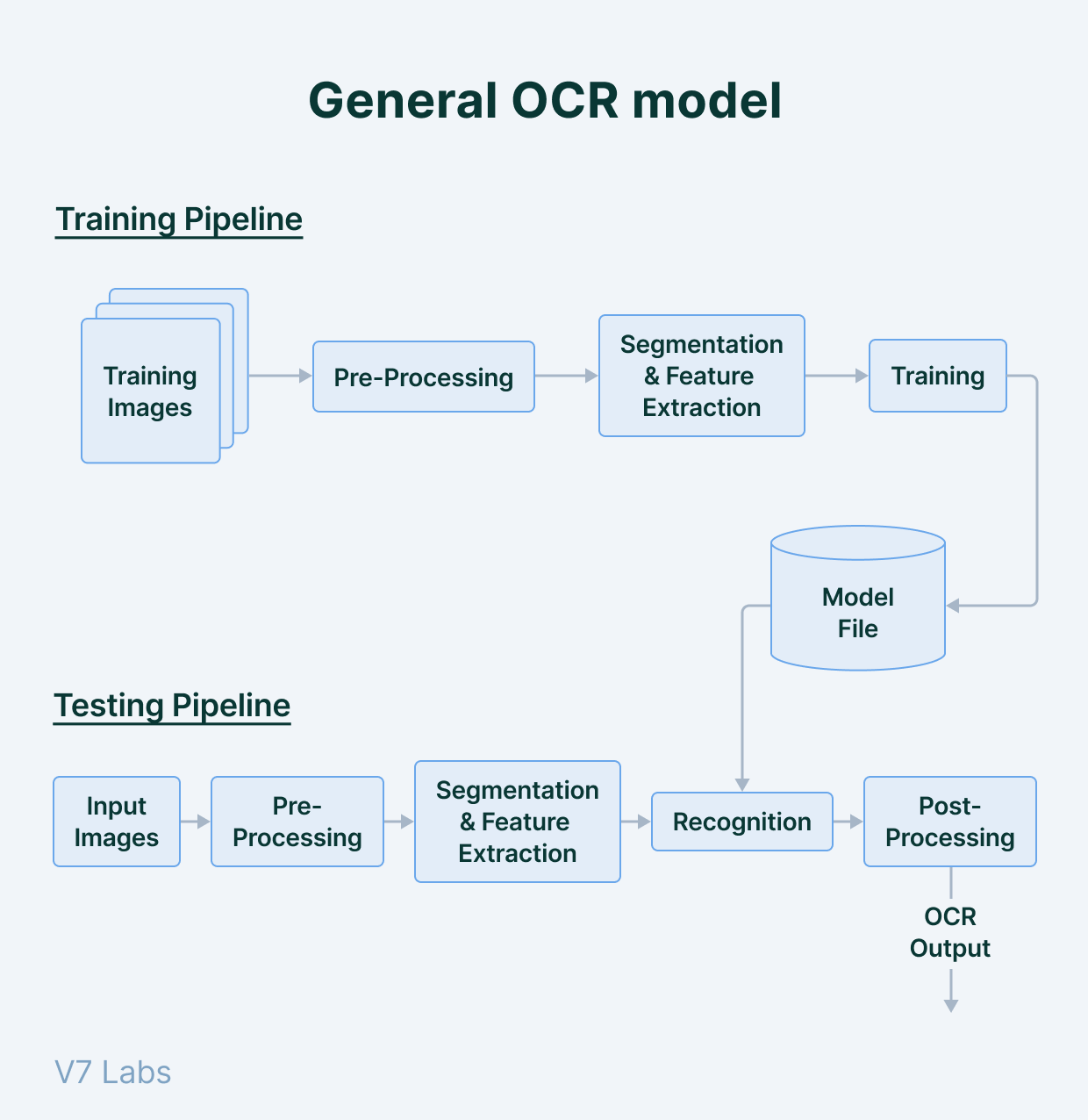
**Báo cáo 11/6/2022**

1. **Tên đề tài đồ án:** tìm hiểu và vận dụng các phương pháp OCR trong bài toán rút trích thông tin từ ảnh
2. **Quá trình OCR rút trích thông tin như thế nào:**



1. **Các bài toán dự định làm:**

* Nhận dạng chữ viết tay (tiếng Việt, tiếng Anh). Tú
* Lấy chữ trong ảnh captcha: [CaptchaOCR](https://docs.google.com/document/d/1jZACHeEwXvuqoD8PlLOXvvKiN4i9vuiMM0lq6MB-WxE/edit?usp=sharing)
* Nhận dạng thông tin (thẻ sinh viên, cmnd,...) qua ảnh, camera [ID CARD OCR](https://docs.google.com/document/d/1FAifMXFqN8bAXeAdkd3ywlDtFvc0CkmNUHZT0SXvxIk/edit?usp=sharing)
* Nhận diện biển số xe thông qua video. Tín [ANPR](https://docs.google.com/document/d/1aK2sivosl5__2AaNmkVok9TADBNiKHaPZpLGdR9ggaw/edit#)
* Tự thực hiện lại mã nguồn một quá trình OCR đơn giản với 1 trong các bài toán trên

1. **Các thư viện OCR tìm hiểu:**

* VietOCR (mạnh cho tiếng Việt) Phú 1
* Tesseract OCR (mạnh cho tiếng Anh): All
* Paddle OCR (tùy chọn)

1. **Các công đoạn tìm hiểu:**

* Kiến trúc model, các thư viện OCR
* Các pipeline, end2end
* Tập data tương ứng
* Preprocessing: xử lý ảnh bị nhòe, làm nổi bật đối tượng cần phát hiện, calibration,.... (OpenCV)
* Xây dựng một baseline
* Đánh giá so sánh mô hình (benchmark) và phát triển mô hình
* **Đóng gói và tạo API hoặc làm một UI để sử dụng mô hình (tùy chọn)**

1. **Các nguồn tìm hiểu:**

* **Paper:** <https://storage.googleapis.com/pub-tools-public-publication-data/pdf/35248.pdf> (Tesseract OCR)

<https://arxiv.org/pdf/2111.15664v4.pdf> (TransformerOCR)

<https://arxiv.org/pdf/2103.10213v1.pdf> (Reciept OCR)

<https://arxiv.org/pdf/2109.03144v2.pdf> (Light-weights OCR)

<https://arxiv.org/pdf/1507.05717v1.pdf> (RCNN)

* **Github:**

<https://github.com/PaddlePaddle/PaddleOCR>

<https://github.com/pbcquoc/vietocr>

<https://github.com/tesseract-ocr/tesseract>

1. **Công việc dự định tuần tới:**

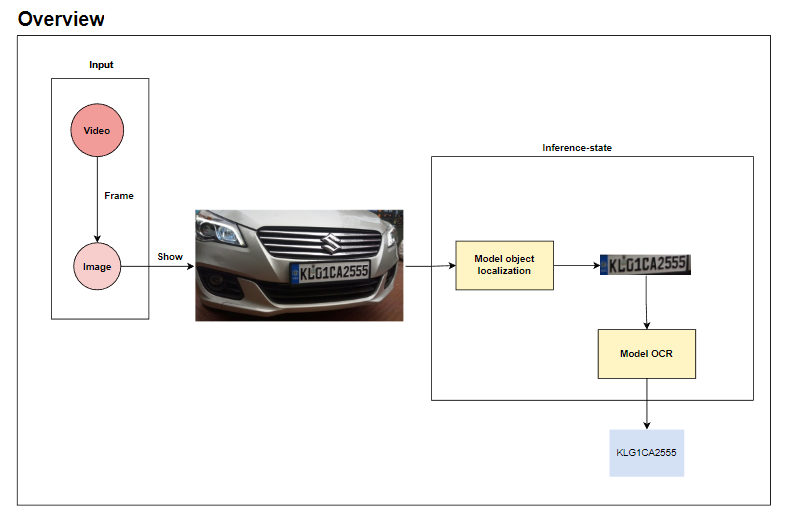
* **Thống nhất đề tài - list công việc - tập dữ liệu**
* Đọc paper và vẽ lại kiến trúc các mô hình, liệt kê điểm yếu điểm mạnh, đặc biệt.
* Một vài tiếp cận trong thao tác tiền xử lý ảnh.
* Thử các pretrain model.
* Bắt đầu xây dựng

**Báo cáo 14/10/2022**

1. **Thống nhất đề tài:**

* OCR với biển số xe.
* Framework chọn: tensorflow

1. **Mô hình:**

****

1. **Dataset:**

* <https://www.kaggle.com/datasets/tolgadincer/us-license-plates> (5000 file - màu)
* <https://www.kaggle.com/datasets/tustunkok/synthetic-turkish-license-plates> (100000 file - đen trắng)
* <https://www.kaggle.com/datasets/francescopettini/license-plate-characters-detection-ocr> (500 file - bounding box phù hợp cho những mô hình CNN)

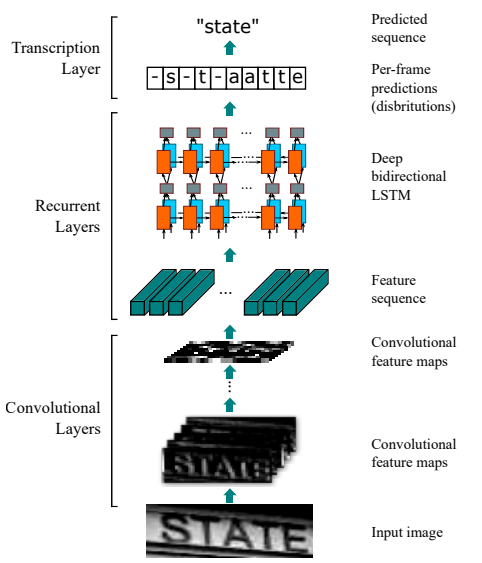
1. **Các mô hình OCR đã tìm hiểu:**

* Mô hình dựa vào retrieval image: [https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.713.886&rep=rep1&type=pdf](https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.713.886&rep=rep1&type=pdf&fbclid=IwAR3mjZr65pFDGIMyfGFcjAv90QI1rphaKcEL8C5g1eL3J4bxTtTRdN-DCkY)
* Mô hình classification:

<https://www.cv-foundation.org/openaccess/content_iccv_2013/papers/Bissacco_PhotoOCR_Reading_Text_2013_ICCV_paper.pdf?fbclid=IwAR3mjZr65pFDGIMyfGFcjAv90QI1rphaKcEL8C5g1eL3J4bxTtTRdN-DCkY>

* Mô hình end-2-end (C-RNN):

<https://arxiv.org/pdf/1507.05717v1.pdf?fbclid=IwAR0pRb7H8L9TaTIfELIU5ED4Y5DXDG0frFsYiGdpS5LycBpXwg_UMsx1z_4>



* Mô hình end-2-end kết hợp xử lý ảnh:

<http://www.bmva.org/bmvc/2016/papers/paper043/paper043.pdf?fbclid=IwAR2r0i7WOM0nsT_3QGEL22OZQ1VEVrJIKJGsTvzzX27eBejE1HOGwUlixLs>

1. **Công việc tuần tới:**

* Customize dataset
* Xây dựng mô hình OCR chạy được và giải thích (theo các hướng ở phía trên).

**Báo cáo 27/10/2022**

1. **Công việc đã làm tuần này:**

* **Thống nhất 2 model để hiện thực:**

+ Model classification + Bbox extractor scripting

+ Model End-2-End CRNN

* **Customize dataset:**

+ Data for classification task:

<https://www.kaggle.com/datasets/tustunkok/synthetic-turkish-license-plates>

+ Data for End2Endtask:

<https://www.kaggle.com/datasets/tustunkok/synthetic-turkish-license-plates>

* **Hiện thực Model classification**

Training: <https://colab.research.google.com/drive/180BQsOT_uke56BBtG9A0sI0pk0ZQP6fv?usp=sharing>

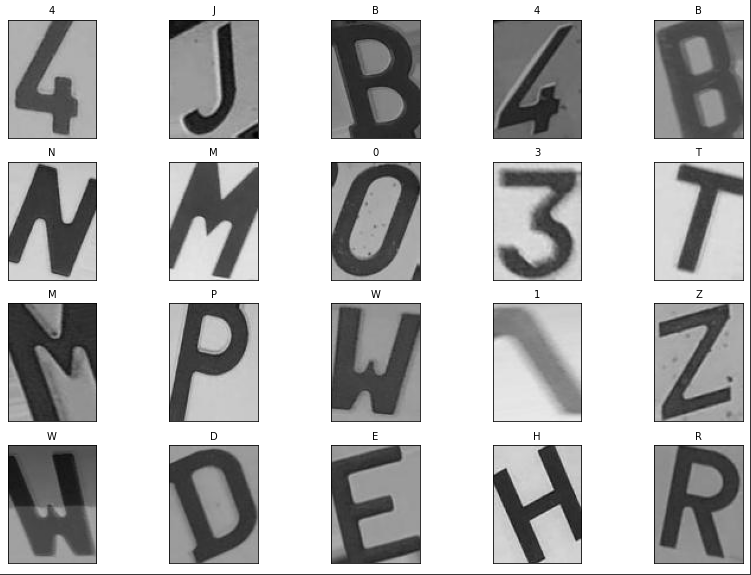
Kết hợp scripting: <https://colab.research.google.com/drive/1XzsCimKaG1Uyvuk2gSCQN4kKSLURym2a?usp=sharing>

* **Tìm hiểu về bắt đầu xây dựng model End2End**
* **Tìm hiểu về model localization và các công nghệ liên quan**

1. **Model classification:**

**Data description:**

Data gồm 100000 sample ảnh kích thước (128,128,3) nhưng đã qua các bước xử lý augmentation và chuyển qua grayscale, mỗi hình ảnh là một chữ riêng biệt và nhãn sẽ là chữ đó. VD:



**Training description:** + Train/val/test: 72/8/20

+ batch\_size: 512

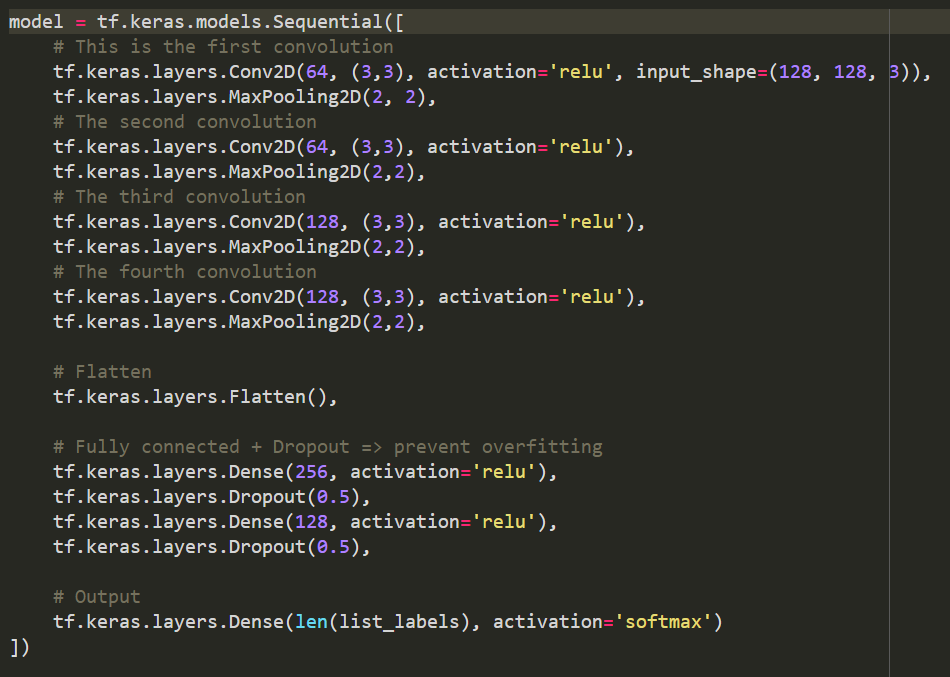
+ epoch: 10

**Model description:**

**+ input:** imageshape(128,128,3)

**+ Output:** label of image

**+ Architect:**

****

**Result:**

+ Time training: 4 tiếng (bật GPU)

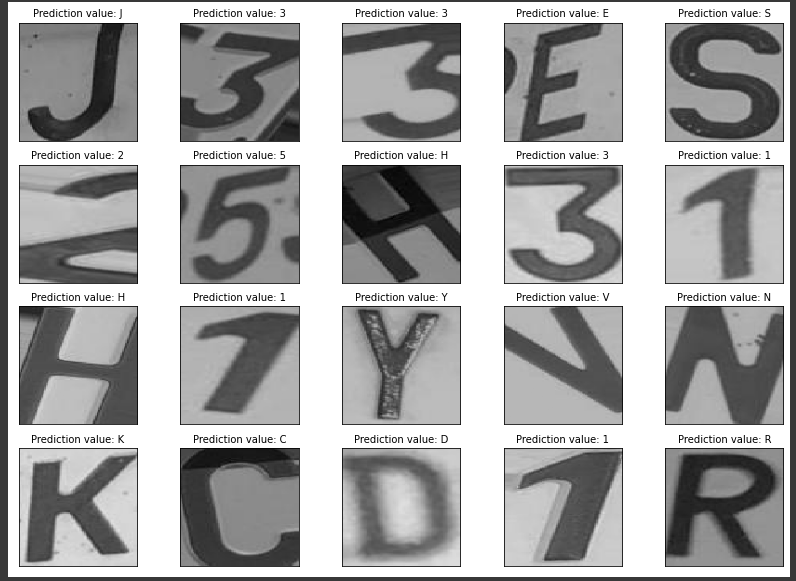
+ acc, loss trên tập train:



+ Độ chính xác trên tập test:

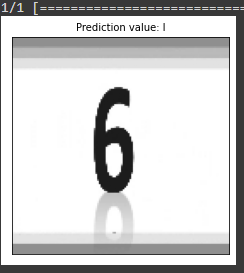
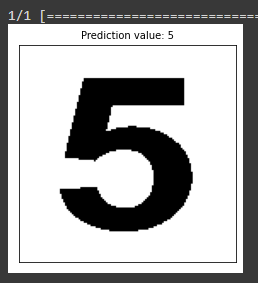
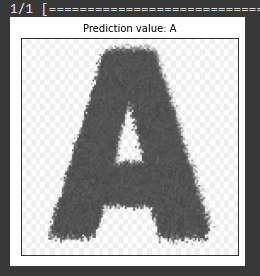


**Một vài sample được dự đoán của tập test:**

****

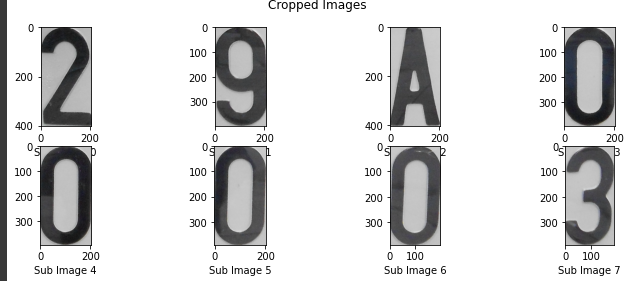
**Một vài sample ở bên ngoài:**

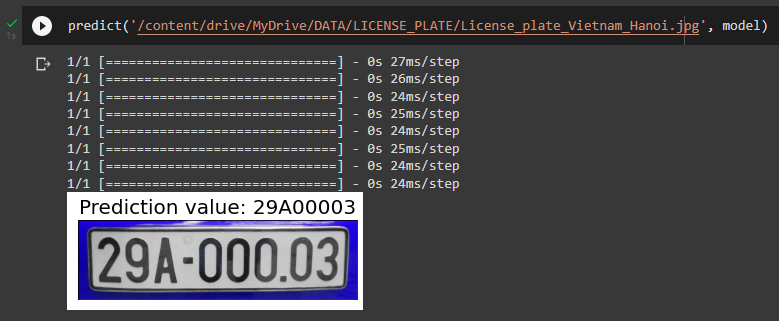
* Ảnh bên ngoài sẽ preprocess lại thành ảnh grayscale với kích thước (128,128,3) phù hợp với model.
* Model adapt khá tốt với nhiều loại data khác, model chỉ sai với dữ liệu có dạng chữ quá nhỏ trong hình nhưng bbox extractor khi trích xuất ra các chữ bự chiếm đa số khung hình nên ta không lo ngại điều này trên tập data biển số xe.

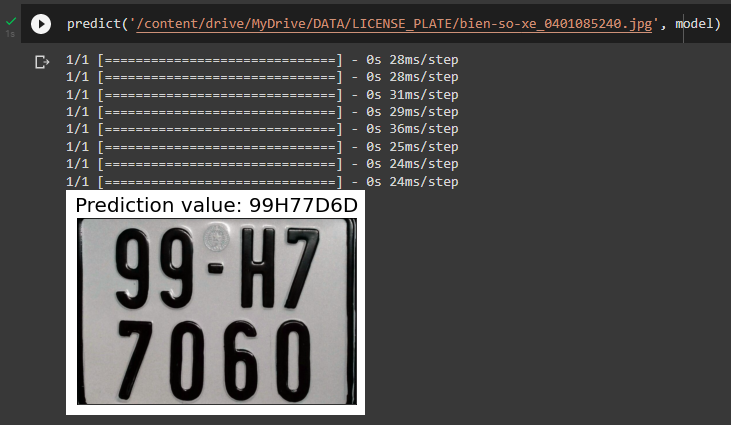


**Kết hợp với bbox extractor:**

* **Bbox extractor** được nhóm viết là một script tự động để trích xuất các chữ trong hình ra dựa vào các thuật toán xử lý ảnh cũng như là thuật toán ad-hoc của nhóm.
* **Có hai kiểu biển số xe chính:** Các chữ nằm trong một dòng và chia ra 2 dòng. Cả hai trường hợp này với tập ảnh kiếm được, đều trích xuất được hết đúng chữ cũng như là đúng thứ tự





**Điểm mạnh:** Thời gian cho ra kết quả nhanh và model khá nhẹ

**Điểm yếu:** do script tự động trích xuất nên vẫn còn một số case sai trích xuất sai mà nhóm chưa phát hiện được.

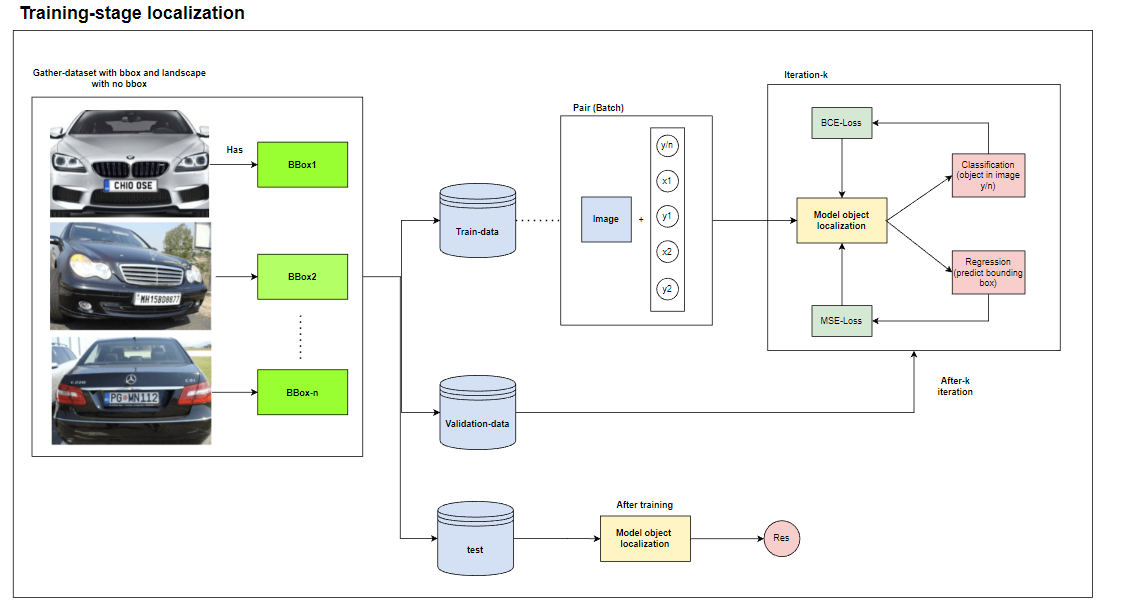
1. **Model localization:**

Input: image + vector description (y/n, coordinate of bounding box)

Output: classification (object in image) && Regression (coordinate prediction of bbox)

Một vài hướng tiếp cận:

* Xây dựng các mô hình SSD, Fast RCNN, RCNN (cần một tập lớn dữ liệu bounding box)
* Config lại Yolo model chỉ quan tâm đến các vật thể bảng số xe.

****

1. **Công việc dự định tuần tới:**

* Đẩy mạnh việc thực hiện model end2end
* Chọn hướng tiếp cận localization (thiên hướng: config lại yolo)
* Thử một vài trường hợp khác cho model classification (ảnh biển số xe bị nghiêng bị xoay, bị mờ).