

گزارش فاین تیون کردن Paddle OCR ۲.۶ بر روی متن فارسی (دست نویس و ...):

[لینک کد قابل اجرا در کولب](#)

برای انجام اینکار از ورژن آخر کتابخانه Paddle OCR موجود در گیت هاب استفاده شده است ، این کتابخانه مراحل ساده ای را برای فاین تیون کردن خودش با داده های جدید که میتواند شامل زبان جدیدی نیز باشد معرفی کرده است، برای شروع مراحل ابتدا کتابخانه مربوطه را از گیت هاب کلون کنید.

- **توجه :** فرایند گفته شده بدلیل عدم امکان نصب پیش نیاز ها و Downgrade,Upgrade آنها در کولب قابل اجرا نیست ، همچنین مراحل گفته شده در ویندوز طی شده اند اما مدل بیشتر از نیم ساعت برای یک ایپاک در حال اجرا بوده و خروجی مبنی بر شروع آموزش یا خطا در روند آن نمایش نمی دهد !
- نتیجتاً فرایند فاین تیون مدل بر روی سیستم عامل مک (و لینوکس) اجرا شده است و لازم است که از ورژن  $\geq 2$  numpy استفاده شود (بر اساس بررسی های زیاد، نسخه سازگار ۱.۲۶ numpy است)

راحتی مراحل انجام فاین تیون کردن از این لینک قابل مشاهده است :

<https://anushsom.medium.com/finetuning-paddleocrs-recognition-model-for-dummies-by-a-dummy-۸۹ac۷d۷edcf۶>

مراحل فاین تیون کردن Paddle OCR بصورت خلاصه شامل این گام هاست :  
داده های نشانه گذاری شده که بصورت دو ستونه در فایل اکسل وجود دارند تهیه می شوند (ستون اول ادرس تصویر و ستون دوم متن معادل آن است)

فایل اکسل مربوطه را به CSV تبدیل میکنیم و دیتاست را به بخش های غیر همپوشان train,valid,test تقسیم بندی میکنیم.(سه پوشه جداگانه بسازید که هر پوشه تنها شامل تصاویر مربوطه باشد)

### نکات مربوطه :

یک فایل txt که شامل الفبای زبان است باید تهیه شود (در هر خط باید یک کاراکتر مشخص شده باشد) ، توجه کنید بخش متنی که در دیتاست نشانه گذاری قرار دارد باید تنهای شامل حروف الفبای فایل دیکشنری زبان باشد در غیر اینصورت هنگام آموزش مدل ارور خواهیم داشت.

۲-در ستون اول آدرس تصویر باید بصورت آدرس مطلق (absolute) باشد و به فاصله یک tab با متن متناظر آن تصویر فاصله داشته باشد.

دیتاست CSV مربوطه باید به شکل فایل txt ایجاد شود برای این منظور از دستور زیر استفاده میکنیم :

```
python gen_label.py --mode="rec" --input_path={path to csv file} --output_label={folder to output txt.file}
```

**توجه:** استفاده از دستور پیشفرض بالا برای تبدیل CSV به txt برای زبان فارسی جواب درستی نمی دهد و فاصله بین کلمات را حذف میکند (مشابه تصویر زیر) و کلمات بهم چسبیده در فایل نهایی قرار می گیرند. برای جلوگیری از این مشکل از اسکریپت موجود در مخزن گیت هاب استفاده کنید که کاما را با tab در فایل جایگزین میکند و فرمت نهایی را ایجاد میکند

| en_PP-OCRv3_rec.yml       | output_label.txt X        |
|---------------------------|---------------------------|
| 1 /content/images/17.jpg  | دروجهآقاییگی              |
| 2 /content/images/18.jpg  | دروجهآجمهدیوسفی           |
| 3 /content/images/19.jpg  | دروجهآجسینصغری            |
| 4 /content/images/20.jpg  | دروجهآجمنعاقورطوبتی       |
| 5 /content/images/21.jpg  | دروجهآقاییگی              |
| 6 /content/images/22.jpg  | دروجهآقاییگی              |
| 7 /content/images/23.jpg  | دروجهآقاییگی              |
| 8 /content/images/24.jpg  | دروجهآقاییجلالی           |
| 9 /content/images/25.jpg  | دروجهآقاییجلالی           |
| 10 /content/images/26.jpg | دروجهآنامابولحسنیباتتسویه |
| 11 /content/images/27.jpg | دروجهآشکرکتکشتیرانیحافظ   |
| 12 /content/images/28.jpg | دروجهآشکرکتکشتیرانیحافظ   |
| 13 /content/images/29.jpg | دروجهآسترگامبهروز         |
| 14 /content/images/30.jpg | دروجهآسترگامبهروز         |

**نکته:** در صورت رعایت نکردن فاصله یک tab بین آدرس مطلق و متن تصویر ، موقع آموزش مدل دیتاست بدرستی تجزیه نمی شود و خطای parse error نمایش داده می شود به همین دلیل بعد از تبدیل خودتان فایل txt را دستی هم چک کنید.(ترجیحا نام فایل هم دارای فاصله هم نباشد بهتر است)

در گام بعدی فایل دیکشنری زبان خودمون را مشخص میکنیم که هر خط آن یک کاراکتر از زبان ما را باید شامل شود و با فرمت txt ذخیره کنیم .

همچنین باید یک مدل pre trained را دانلود کنیم تا از وزن ها و معماری آن برای تسک جدید استفاده کنیم برای مثال برای فاین تیون کردن بر روی داده های انگلیسی از مدل en\_ppocr\_v3\_rec model استفاده می کنیم که قابل دانلود از سایت رسمی Paddle OCR است.([این لینک](#)) .

همچنین برای زبان فارسی در اینجا از وزن های نسخه fa\_mobile\_ver2.0\_rec برای آموزش مدل استفاده شده است

**فایل کانفیگ yml:** فایل با فرمت yml دانلود شده با مخزن اصلی Paddle OCR وجود دارد که شامل پارامترهای کانفیگ مدل است، این فایل فیلدهای زیادی دارد که باید بر اساس ادرسهای سیستم شما و هایپر پارامترهایی که مدنظر دارید مقداردهی شوند:

بخشهای مهم این فایل: امکان استفاده از gpu، مسیر ذخیره نتایج مدل، مسیر دیکشنری زبان مدنظر، فایل لاگ ذخیره روند آموزش مدل، تعداد اپیکهای مدل، تعداد batch size مدل، ادرس پوشه train و فایل txt مربوط به نشانه گذاری تصاویر پوشه train، ادرس پوشه valid و فایل txt مربوط به نشانه گذاری تصاویر پوشه valid، پارامترهای مربوط به آموزش و بهینه سازی مدل (optimizer, regularizer, le) و ...

➔ **توجه:** تعداد batch-size نباید از اندازه دیتاست بیشتر باشد در غیر اینصورت ارور داریم. همچنین با بزرگ کردن این هایپر پارامتر مقدار حافظه رم لازم برای آموزش مدل افزایش می یابد.

```
1 Global:
2   debug: false
3   use_gpu: false # Set to true if you have a GPU available
4   epoch_num: 1
5   log_smooth_window: 20
6   print_batch_step: 2
7   save_model_dir: /Users/hmi/Documents/paddle/Paddle
8   save_epoch_step: 1
9   eval_batch_step: [0, 2]
10  cal_metric_during_train: true
11  pretrained_model: "" # Provide a valid path if you have a pretrained model
12  checkpoints: "" # Provide a valid path if you have checkpoints
13  save_inference_dir: ""
14  use_visualdl: false
15  infer_img: /Users/hmi/Documents/paddle/Paddle/test/1.jpg
16  character_dict_path: /Users/hmi/Documents/paddle/PaddleOCR-main/ppocr/utils/en_dict.txt
17  max_text_length: &max_text_length 25
18  infer_mode: false
19  use_space_char: true
20  distributed: true
21  save_res_path: /Users/hmi/Documents/paddle/Paddle/predicts_ppocrv3_en.txt
22
23 Optimizer:
24   name: Adam
25   beta1: 0.9
26   beta2: 0.999
```

\***توجه:** هایپر پارامترهایی که در این فایل قرار گرفته اند در ادامه روند آموزش مدل بر اساس فاکتورهایی که مشخص کرده ایم ممکن است کاهش یا افزایش داشته باشند (مثلا در صورت وجود Plateau در روند آموزش مدل مقدار LR ب نرخ مشخصی کاهش می یابد).

در نهایت بعد از تقسیم بندی دیتاست و ایجاد فایل txt برای نشانه گذاری هر بخش، مشخص کردن دیکشنری زبان، دانلود مدل های پیش یادگرفته و تنظیم پارامترهای فایل YAML نوبت به اجرای فایل train.py موجود در کتابخانه Paddle است.

```
Python3 tools/train.py -c {path to config file} -o Global.pretrained_model={path to pretrained model}/best_accuracy
```

بعد از اتمام ایپاک های مشخص شده برای آموزش مدل ، وزن های مدل و لاگ آموزش مدل در مسیر های مشخصی شده که در فایل yml وجود داشت ذخیره می شود. با اجرای دستور بالا معمولا وزن های آخرین مدل و بهترین مدل (از نظر معیار accuracy) و همچنین یک فایل yml در پوشه مربوطه ذخیره می شوند.

بعد از اتمام آموزش مدل ، با دستور زیر مدل استنتاج inference model را نیز بر اساس مدل آموزش دیده ایجاد میکنیم:

```
python3 tools/export_model.py -c {path to yml file inside the fine tuned model folder} -o Global.pretrained_model={path to model folder} Global.save_inference_dir={path to inference model folder}
```

با اجرای این دستور خروجی از مدل که برای پیش بینی و اعمال OCR بر روی تصویر بکار می رود ساخته می شود **نکته:** دقت کنید که فایل yml در این دستور فایلی نیست که در ابتدا پیش از آموزش مدل خودمان ایجاد و ویرایش کردیم بلکه بعد از اتمام آموزش در پوشه خروجی که وزن ها قرار دارد یک فایل با نام config.yml ایجاد می شود و باید این فایل به عنوان ورودی برای inference داده شود.

**\*توجه:** نکته قابل توجه در این قسمت استفاده مشترک Paddle OCR و DocTR از معماری MobileNet , Parseq است که در واقع نشان دهنده دلیل موفقیت این دو کتبخانه نیز است.(با این تفاوت که یکی از MobileNetV3 و دیگری از MobileNetV1 enhanced استفاده می کنند)