حذف مهر ها از چک : (لینک گیت هاب)

بعد از اعمال پایپلاین پیش پردازش تصاویر ، تصاویر در پلتفرم Roboflow نشانه گذاری شدند و مهر های مربوطه به هر چک مشخص شد ، بعد از اماده سازی دیتاست و اعمال داده افزایی های مختلف روی ان (blurring,noisy,rotation,grayscale , ...) در نهایت با مدل YOLOv۸X تسک تشخیص شی مهر بر روی داده ها اعمال شد.(آموزش مدل با ۴۰ ایپاک انجام شد)

(لینک عملکرد مدل YOLOv۸X آموزش داده شده)

نمونه مهر تشخیص داده شده Yolov۸ و conf=۰.۸۳

در نهایت مدل آموزش دیده را بر روی تمام تصایر چک ها اعمال کردیم تا مختصات مهر های موجود در چک را مشخص کند (با conf=۰.۴) بعد از تشخیص انها کدی که با استفاده از تابع inpaint از کتابخانه Opencv بر روی تصاویر اعمال شد و مهر ها از تصویر حذف شدند (لینک خروجی های اصلاح شده)

نمونه خروجی ها :

در نمونه اخر بدلیل اینکه محدوده bounding box توسط YOLO کامل مشخص نشده است کل مهر از تصویر حذف نشده است.

پیشنهاد : برای مواردی که میخواهیم مهر حذف شود اما به متن و دیتا چک اسیبی نرسد بهتر است از OCR هم برای حذف استفاده کنیم هرچند حالت بهتر استفاده از روش های یادگیری عمیق و شبکه GAN است که ای موارد دیتاست و ماسک متناظر آنها را در مقیاس زیاد می طلبد.

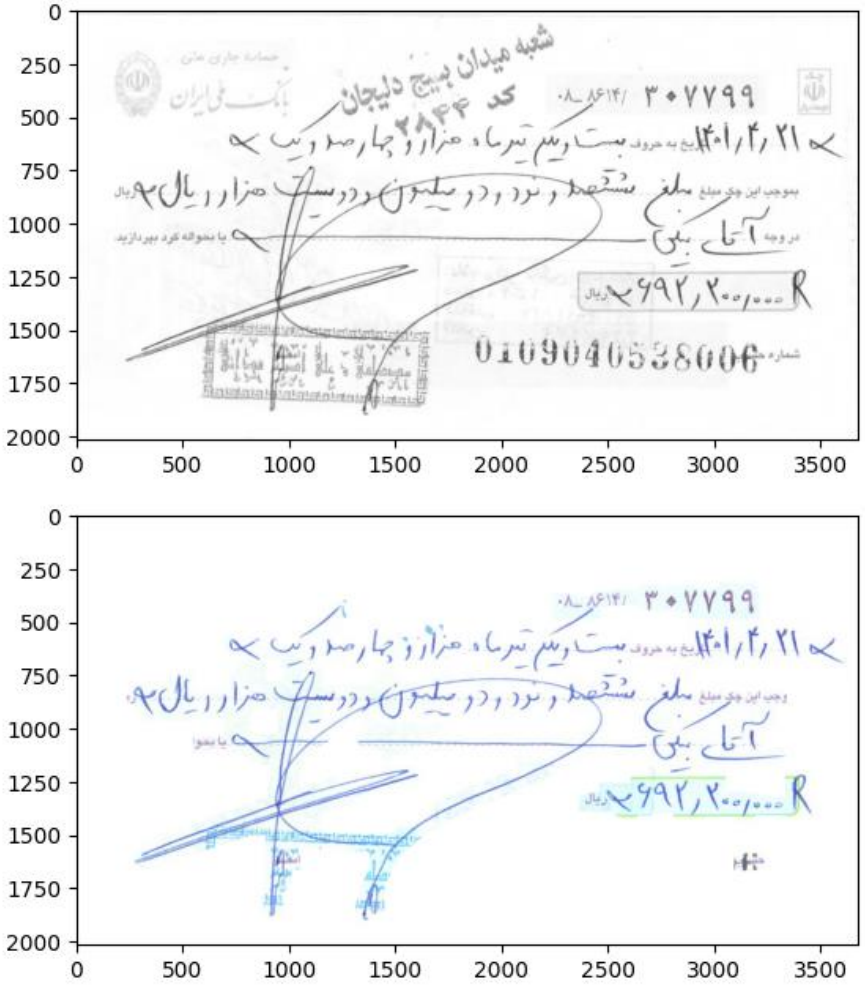
➔ نمونه مخزن گیت هاب حذف واتر مارک با GAN , Inpainting

روش دوم (مبتنی بر بافت تصویر و خوشه بندی) : در این روش از الگوریتم خوشه بندی agglomerative استفاده شده است که تصویر طی چند مرحله به نما های گوناگون تبدیل شده و خروجی نهایی که حاصیل تفاضل مهر استخراج شده و تصویر اصلی است را ذخیره می کنیم .

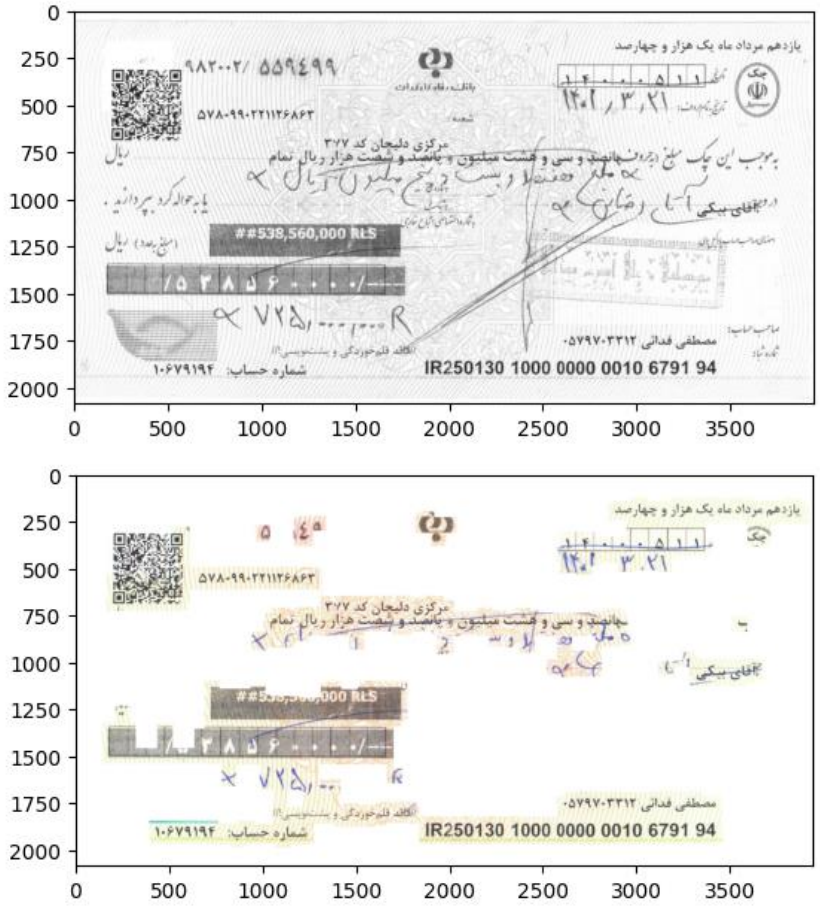
ابتدا روی تصویر یک آستانه گیر ادپتیو اعمال می کنیم ، سپس تصویر را معکوس می کنیم تا نوشته ها بخوبی تشخیص داده شوند ، بعد از آن عملگر مورفولوژیک dilation اعمال می کنید تا متون استخراج کرده را به هم متصل کند تا پیوستگی حروف حفظ شده و در حذف مهر بافت متنی آسیب نبیند.

در خورجی های مختلفی که کد تولید می کند خروجی final همان تصویری ز چک است که مهر از روی آن حذف شده است.

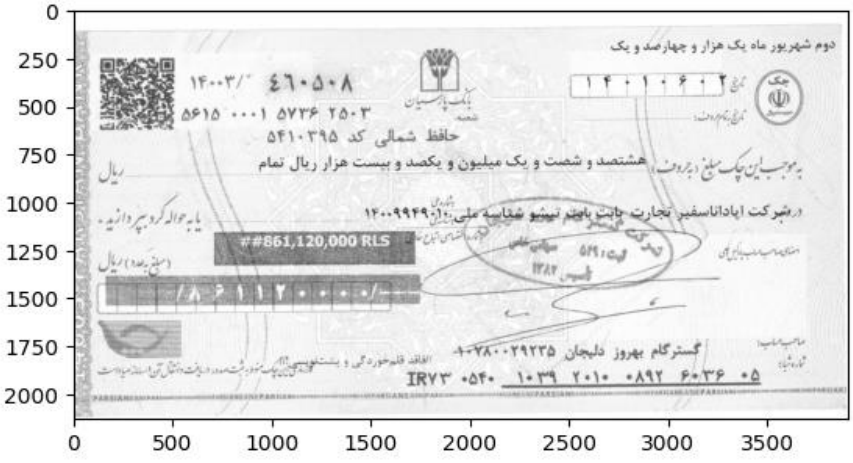
چند نمونه ورودی و خروجی این روش :

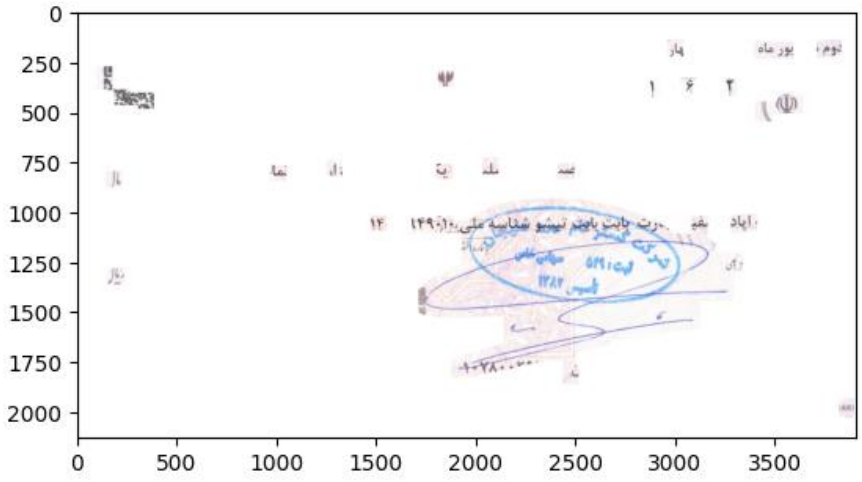


ورودی خروجی شماره ۱



ورودی خروجی شماره ۲





ورودی و خروجی شماره ۳ (خوشه بندی اشتباه همراه با از دست دادن دیتا)

لینک نتایج اعمال شده با این روش

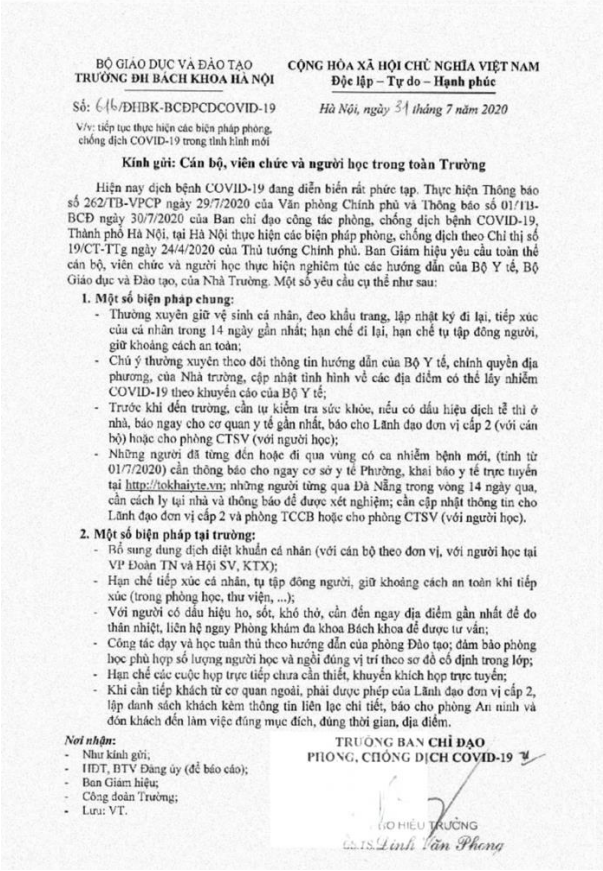
روش ترکیبی : اعمال easy ocr برای حذف بدون از دادن داده :

با ترکیبب گفته شده در بالا با کتابخانه ocr برای زبان فارسی و انگلیسی بعد از تشخیص مهر و ناحیه هایی که باید حذف شوند ابتدا یک OCR بر روی تصویر اعمال می شود تا از پاک کردن متن بخش های مختلف چک جلوگیری کند ، هر چند عملکرد این روش چندان خوب نبود چرا که خود مهر ها دارای متن هم هستند و باعث میشد مهر به خوبی حذف نشود اما میتواند درجه حساسیت مدل را کمتر کرد تا به این بخش ها حساسیت کمتری داشته باشد.

بررسی کتابخانه های متن باز حذف مهر

https://github.com/sun-asterisk-research/stamp_processing?tab=readme-ov-file

این مخزن شامل مدل یادگیری عمیق آموزش دیده روی تعداد زیادی سند است و در حذف مهر با حفظ بکگراند آن عملکرد خوبی دارد ، این مدل دارای دو فاز است ، فاز اول تشخیص مهر با مدل yolov۵ و در فاز بعدی محدوده مهر را حذف میکند، نمونه تصویر از حذف مهر توسط این مدل:



BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐH BÁCH KHOA HÀ NỘI

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập – Tự do – Hạnh phúc

Số: 616/ĐHBK-BCĐPCDCOVID-19

Hà Nội, ngày 31 tháng 7 năm 2020

V/v: tiếp tục thực hiện các biện pháp phòng, chống dịch COVID-19 trong tình hình mới

Kính gửi: Cán bộ, viên chức và người học trong toàn Trường

Hiện nay dịch bệnh COVID-19 đang diễn biến rất phức tạp. Thực hiện Thông báo số 262/TB-VPCP ngày 29/7/2020 của Văn phòng Chính phủ và Thông báo số 01/TB-BCĐ ngày 30/7/2020 của Ban chỉ đạo công tác phòng, chống dịch bệnh COVID-19, Thành phố Hà Nội, tại Hà Nội thực hiện các biện pháp phòng, chống dịch theo Chỉ thị số 19/CT-TTg ngày 24/4/2020 của Thủ tướng Chính phủ. Ban Giám hiệu yêu cầu toàn thể cán bộ, viên chức và người học thực hiện nghiêm túc các hướng dẫn của Bộ Y tế, Bộ Giáo dục và Đào tạo, của Nhà Trường. Một số yêu cầu cụ thể như sau:

1. Một số biện pháp chung:

- Thường xuyên giữ vệ sinh cá nhân, đeo khẩu trang, lập nhật ký đi lại, tiếp xúc của cá nhân trong 14 ngày gần nhất; hạn chế đi lại, hạn chế tụ tập đông người, giữ khoảng cách an toàn;
- Chú ý thường xuyên theo dõi thông tin hướng dẫn của Bộ Y tế, chính quyền địa phương, của Nhà trường, cập nhật tình hình về các địa điểm có thể lây nhiễm COVID-19 theo khuyến cáo của Bộ Y tế;
- Trước khi đến trường, cần tự kiểm tra sức khỏe, nếu có dấu hiệu dịch tễ thì ở nhà, báo ngay cho cơ quan y tế gần nhất, báo cho Lãnh đạo đơn vị cấp 2 (với cán bộ) hoặc cho phòng CTSV (với người học);
- Những người đã từng đến hoặc đi qua vùng có ca nhiễm bệnh mới, (tính từ 01/7/2020) cần thông báo cho ngay cơ sở y tế Phường, khai báo y tế trực tuyến tại http://tokhaiyte.vn; những người từng qua Đà Nẵng trong vòng 14 ngày qua, cần cách ly tại nhà và thông báo để được xét nghiệm; cần cập nhật thông tin cho Lãnh đạo đơn vị cấp 2 và phòng TCCB hoặc cho phòng CTSV (với người học).

2. Một số biện pháp tại trường:

- Bổ sung dung dịch diệt khuẩn cá nhân (với cán bộ theo đơn vị, với người học tại VP Đoàn TN và Hội SV, KTX);
- Hạn chế tiếp xúc cá nhân, tụ tập đông người, giữ khoảng cách an toàn khi tiếp xúc (trong phòng học, thư viện, ...);
- Với người có dấu hiệu ho, sốt, khó thở, cần đến ngay địa điểm gần nhất để đo thân nhiệt, liên hệ ngay Phòng khám đa khoa Bách khoa để được tư vấn;
- Công tác dạy và học tuân thủ theo hướng dẫn của phòng Đào tạo; đảm bảo phòng học phù hợp số lượng người học và ngồi đúng vị trí theo sơ đồ cố định trong lớp;
- Hạn chế các cuộc họp trực tiếp chưa cần thiết, khuyến khích họp trực tuyến;
- Khi cần tiếp khách từ cơ quan ngoài, phải được phép của Lãnh đạo đơn vị cấp 2, lập danh sách khách kèm thông tin liên lạc chi tiết, báo cho phòng An ninh và đón khách đến làm việc đúng mục đích, đúng thời gian, địa điểm.

Nơi nhận:
- Như kính gửi;
- HĐT, BTV Đảng ủy (để báo cáo);
- Ban Giám hiệu;
- Công đoàn Trường;
- Lưu: VT.

TRƯỞNG BAN CHỈ ĐẠO
PHÒNG, CHỐNG DỊCH COVID-19

PGS.TS. Đinh Văn Phong

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐH BÁCH KHOA HÀ NỘI

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập – Tự do – Hạnh phúc

Số: 616/ĐHBK-BCĐPCDCOVID-19

Hà Nội, ngày 31 tháng 7 năm 2020

V/v: tiếp tục thực hiện các biện pháp phòng, chống dịch COVID-19 trong tình hình mới

Kính gửi: Cán bộ, viên chức và người học trong toàn Trường

Hiện nay dịch bệnh COVID-19 đang diễn biến rất phức tạp. Thực hiện Thông báo số 262/TB-VPCP ngày 29/7/2020 của Văn phòng Chính phủ và Thông báo số 01/TB-BCĐ ngày 30/7/2020 của Ban chỉ đạo công tác phòng, chống dịch bệnh COVID-19, Thành phố Hà Nội, tại Hà Nội thực hiện các biện pháp phòng, chống dịch theo Chỉ thị số 19/CT-TTg ngày 24/4/2020 của Thủ tướng Chính phủ. Ban Giám hiệu yêu cầu toàn thể cán bộ, viên chức và người học thực hiện nghiêm túc các hướng dẫn của Bộ Y tế, Bộ Giáo dục và Đào tạo, của Nhà Trường. Một số yêu cầu cụ thể như sau:

1. Một số biện pháp chung:

- Thường xuyên giữ vệ sinh cá nhân, đeo khẩu trang, lập nhật ký đi lại, tiếp xúc của cá nhân trong 14 ngày gần nhất; hạn chế đi lại, hạn chế tụ tập đông người, giữ khoảng cách an toàn;
- Chú ý thường xuyên theo dõi thông tin hướng dẫn của Bộ Y tế, chính quyền địa phương, của Nhà trường, cập nhật tình hình về các địa điểm có thể lây nhiễm COVID-19 theo khuyến cáo của Bộ Y tế;
- Trước khi đến trường, cần tự kiểm tra sức khỏe, nếu có dấu hiệu dịch tễ thì ở nhà, báo ngay cho cơ quan y tế gần nhất, báo cho Lãnh đạo đơn vị cấp 2 (với cán bộ) hoặc cho phòng CTSV (với người học);
- Những người đã từng đến hoặc đi qua vùng có ca nhiễm bệnh mới, (tính từ 01/7/2020) cần thông báo cho ngay cơ sở y tế Phường, khai báo y tế trực tuyến tại http://tokhaiyte.vn; những người từng qua Đà Nẵng trong vòng 14 ngày qua, cần cách ly tại nhà và thông báo để được xét nghiệm; cần cập nhật thông tin cho Lãnh đạo đơn vị cấp 2 và phòng TCCB hoặc cho phòng CTSV (với người học).

2. Một số biện pháp tại trường:

- Bổ sung dung dịch diệt khuẩn cá nhân (với cán bộ theo đơn vị, với người học tại VP Đoàn TN và Hội SV, KTX);
- Hạn chế tiếp xúc cá nhân, tụ tập đông người, giữ khoảng cách an toàn khi tiếp xúc (trong phòng học, thư viện, ...);
- Với người có dấu hiệu ho, sốt, khó thở, cần đến ngay địa điểm gần nhất để đo thân nhiệt, liên hệ ngay Phòng khám đa khoa Bách khoa để được tư vấn;
- Công tác dạy và học tuân thủ theo hướng dẫn của phòng Đào tạo; đảm bảo phòng học phù hợp số lượng người học và ngồi đúng vị trí theo sơ đồ cố định trong lớp;
- Hạn chế các cuộc họp trực tiếp chưa cần thiết, khuyến khích họp trực tuyến;
- Khi cần tiếp khách từ cơ quan ngoài, phải được phép của Lãnh đạo đơn vị cấp 2, lập danh sách khách kèm thông tin liên lạc chi tiết, báo cho phòng An ninh và đón khách đến làm việc đúng mục đích, đúng thời gian, địa điểm.

Nơi nhận:
- Như kính gửi;
- HĐT, BTV Đảng ủy (để báo cáo);
- Ban Giám hiệu;
- Công đoàn Trường;
- Lưu: VT.

TRƯỞNG BAN CHỈ ĐẠO
PHÒNG, CHỐNG DỊCH COVID-19

TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI

PGS.TS. Đinh Văn Phong

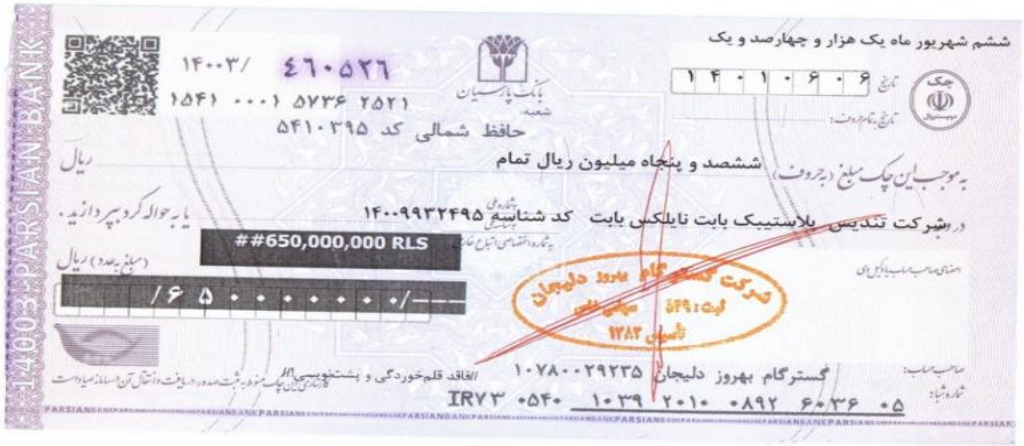
این کتابخانه علی رغم موجود بودن در pypi امکان نصب از طریق دستورpip را ندارد چون نیازمندی pytorch==۱.۸ دارد که گوگل کولب جزو ورژن های ساپورت شده برای نصب آن را ندارد.

خطای نصب دستی ورژن pytorch مناسب این کتابخانه در کولب:

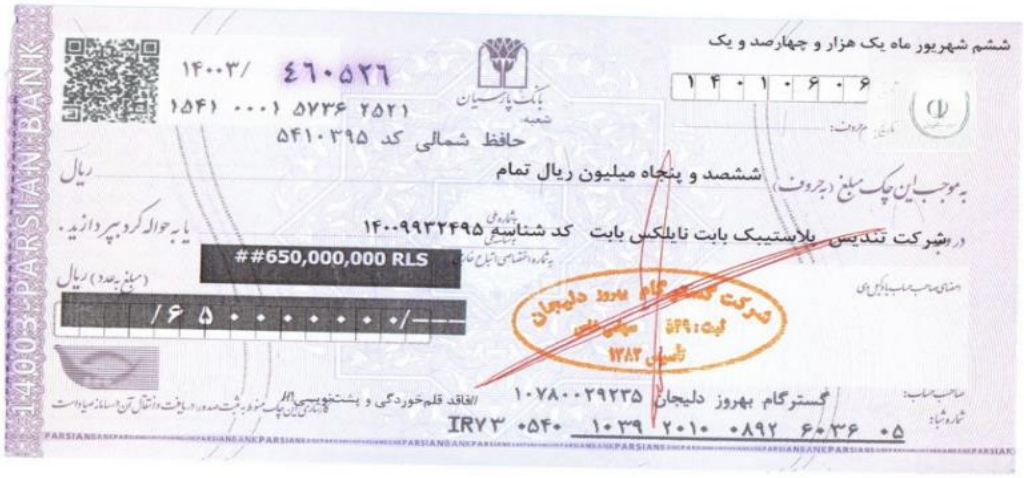
```
ERROR: torch-۱.۸.۱+cu۱۱۱-cp۳۷-cp۳۷m-linux_x۸۶_۶۴.whl is not a supported wheel on this platform.
```

بالاخره برای نصب ان از دستور زیر استفاده شد :

نمونه ورودی :



نمونه خروجی :



همانطور که در این مثال مشخص است مهر حذف نشده و بخشی از ارم بانک بالا سمت راست حذف شده است ، این نتیجه طبیعی است چرا که مدل و وزن های ان بر روی داده های چک اموزش ندیده است

کتابخانه Watermark-Removal: (لینک گیت هاب)

این کتابخانه برای حذف واتر مارک استفاده می شود که نتایج آن بسیار خوب است :

نکته مهم در استفاده از این کتابخانه اجبار در استفاده از تنسورفلو نسخه ۱.۱۵.۰ است در غیر اینصورت اجرای کد با خطا های مختلفی مواجه می شود، بدلیل اینکه کولب تنها از ورژن های بالای ۲ تنسورفلو پشتیبانی میکند و اجازه نصب ورژن کمتر از ۲ را هم نمی دهد (چه با دستور pip install چه از طریق نصب فایل whl کتابخانه) امکان استفاده از آن روی کولب وجود ندارد ...

ارور داده شده برای نصب TF ورژن ۱.۱۵ :

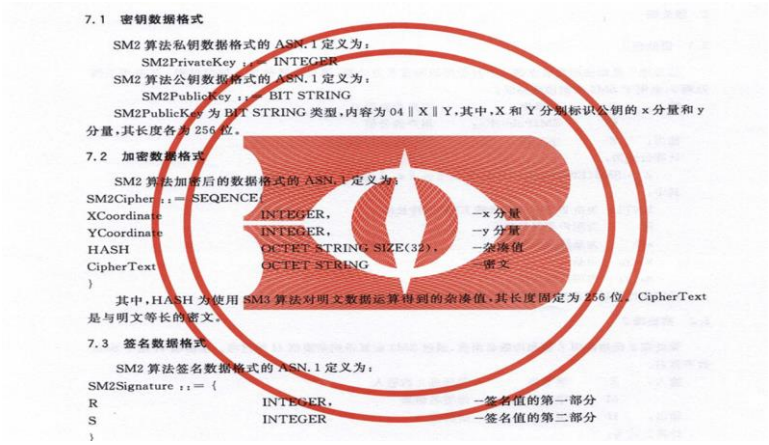
ERROR: tensorflow-۱.۱۵.۰-cp۳۷-cp۳۷m-manylinux۲۰۱۰_x۸۶_۶۴.whl is not a supported wheel on this platform.

سایر کتابخانه ها :

کتابخانه حذف مهر نوشته شده با زبان ruby :این کتابخانه قابلیت نصب و اجرا روی کولب را نداشت

https://github.com/despeck/despeck

ورودی :



7.1 密钥数据格式

SM2 算法私钥数据格式的 ASN.1 定义为：

SM2PrivateKey ::= INTEGER

SM2 算法公钥数据格式的 ASN.1 定义为：

SM2PublicKey ::= BIT STRING

SM2PublicKey 为 BIT STRING 类型，内容为 04 ‖ X ‖ Y，其中，X 和 Y 分别标识公钥的 x 分量和 y 分量，其长度各为 256 位。

7.2 加密数据格式

SM2 算法加密后的数据格式的 ASN.1 定义为：

SM2Cipher ::= SEQENCE{

XCoordinate	INTEGER,	--x 分量
YCoordinate	INTEGER,	--y 分量
HASH	OCTET STRING SIZE(32),	--杂凑值
CipherText	OCTET STRING	--密文

}

其中，HASH 为使用 SM3 算法对明文数据运算得到的杂凑值，其长度固定为 256 位。CipherText 是与明文等长的密文。

7.3 签名数据格式

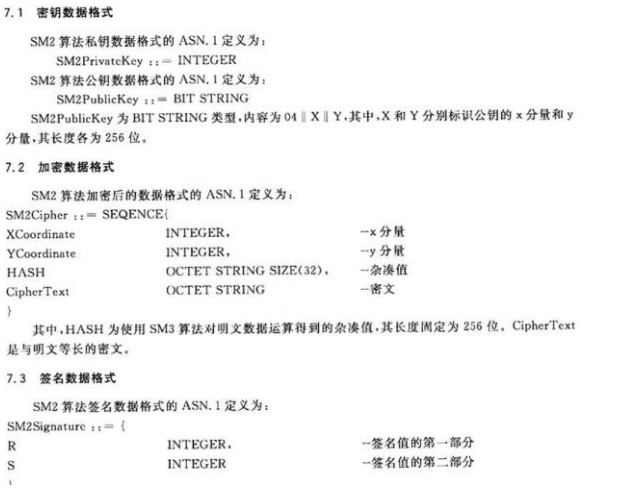
SM2 算法签名数据格式的 ASN.1 定义为：

SM2Signature ::= {

R	INTEGER,	--签名值的第一部分
S	INTEGER	--签名值的第二部分

}

خروجی :



7.1 密钥数据格式

SM2 算法私钥数据格式的 ASN.1 定义为：

SM2PrivateKey ::= INTEGER

SM2 算法公钥数据格式的 ASN.1 定义为：

SM2PublicKey ::= BIT STRING

SM2PublicKey 为 BIT STRING 类型，内容为 04 ‖ X ‖ Y，其中，X 和 Y 分别标识公钥的 x 分量和 y 分量，其长度各为 256 位。

7.2 加密数据格式

SM2 算法加密后的数据格式的 ASN.1 定义为：

SM2Cipher ::= SEQENCE{

XCoordinate	INTEGER,	--x 分量
YCoordinate	INTEGER,	--y 分量
HASH	OCTET STRING SIZE(32),	--杂凑值
CipherText	OCTET STRING	--密文

}

其中，HASH 为使用 SM3 算法对明文数据运算得到的杂凑值，其长度固定为 256 位。CipherText 是与明文等长的密文。

7.3 签名数据格式

SM2 算法签名数据格式的 ASN.1 定义为：

SM2Signature ::= {

R	INTEGER,	--签名值的第一部分
S	INTEGER	--签名值的第二部分

}

این مدل پارامتر های زیادی را برای حذف مهر دارد :

- --color ۰۰FF۰۰ – to say watermark is ~ green.
- --sensitivity ۱۲۰ – increases sensitivity (if with default ۱۰۰ watermark is still visible).
- --black-const -۶۰ – by default, Despeck tries to improve text quality by increasing black by -۱۱۰. This may be too much for you, so you can reduce that number.
- --add-contrast – disabled by default, increases output image's contrast.
- --accurate – disabled by default. Applies filters to the area with watermark only, preserving the rest of the image untouched.

مدل مبتنی بر زبان c++ : این مدل مقاله هم شده است ولی از این مدل اجرا نگرفتم و نمونه ورودی و خروجی هم در مخزن ندارد.
https://github.com/soumyadeepdey/Color-Stamp-Removal