

بررسی نقاط قوت و ضعف مدل های حذف پس زمینه و بهبود آنها :

این گزارشکار بر اساس ۱۲ تصویر تست داده شده به مدل ها نوشته شده است که این دیتاست ها اسناد فارسی ، انگلیسی بوده که بعضا کج اسکن شده اند (skewness دارند) ، دست نویس هستند و رزولوشن تصویر و کیفیت نوشتار آنها در بعضی از تصاویر اصلا خوب نیست ... که میتواند سنگ محک مناسبی برای بررسی قابلیت های مدل ها باشد.

مدل easy ocr : هر چند این مدل جزو قوی ترین هاست و برای OCR متن استفاده می شود اما در کاربرد حذف بکگراند و سگمنت کردن متن هم بسیار خوب عمل کرده است ، سگمنت کردن متن از بکگراند را کاملا بر اساس OCR انجام می دهد یعنی متن هایی که از هم جدا هستند را نیز تشخیص می دهد، در مثال های داده شده که تصویر انسان ها وجود داشته است این مدل آن را هم سگمنت می کند (که باید در پس پردازش در مدل های تشخیص چهره حذف شوند)
نقاط قوت :

+ این مدل دقتش را در سگمنتیشن بر روی متن فارسی و انگلیسی حفظ می کند

+ با فاین تیون کردن و یا ترکیب با سایر مدل ها میتواند بسیار بهتر عمل کند.

+ چون این مدل در پیش پردازش خودش skewness و .. را برطرف میکند خروجی این مدل ها تغییر زاویه موثری دارد.



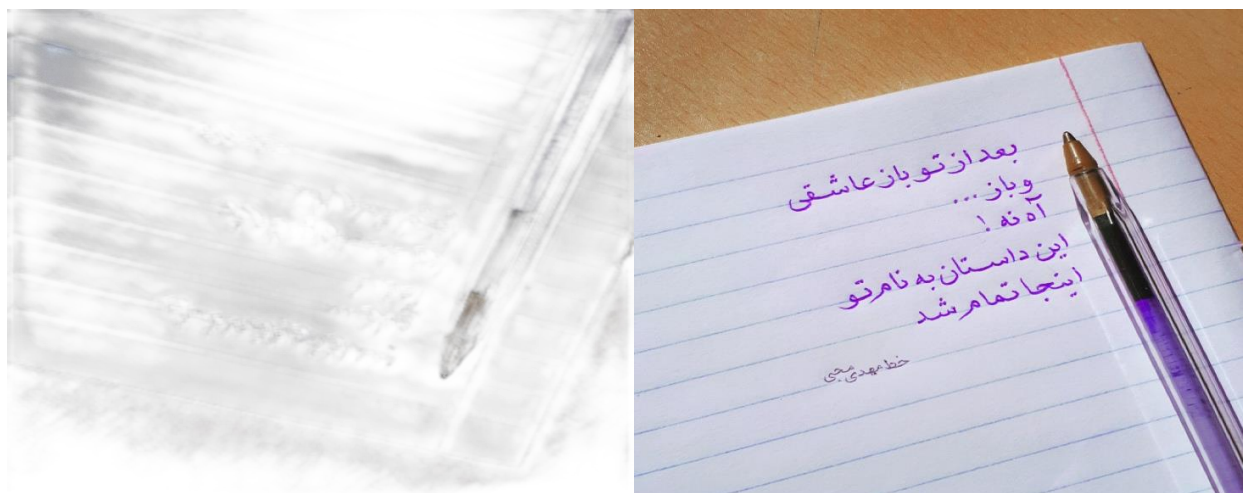
شماره	413	هست فود	23:05:32
ترک دُور	98/10/10		
نام	بی	لعداد	قیمت کل
پیتزا مخلوط مرغ و گوشت	33500	1	33,500
پیتزا استیک بوقلمون	35000	1	35,000
سیب زمینی اسپیشیال	25000	1	25,000
مالشعیر استوایی	4000	1	4,000
مالشعیر کلاسیک	4000	3	12,000
جمع کل :			109,500
قابل پرداخت :			109,500 تومان
ابتدای مطهری پ ۳۷۹	تلفن: ۸۸۱۰۰۰۰۸		

شماره	413	هست فود	23:05:32
ترک دُور	98/10/10		
نام	بی	لعداد	قیمت کل
پیتزا مخلوط مرغ و گوشت	33500	1	33,500
پیتزا استیک بوقلمون	35000	1	35,000
سیب زمینی اسپیشیال	25000	1	25,000
مالشعیر استوایی	4000	1	4,000
مالشعیر کلاسیک	4000	3	12,000
جمع کل :			109,500
قابل پرداخت :			109,500 تومان
ابتدای مطهری پ ۳۷۹	تلفن: ۸۸۱۰۰۰۰۸		

نقاط ضعف: این مدل چون صرفاً بر اساس اسکن متن کار میکند در جایی از تصویر یک حرف الفبا هم ببیند سعی میکند آن را هم شامل شود . (این مورد باید با تنظیم پارامتر beam search تغییر کند یا مراحل پس پردازش اضافی استفاده کنیم .

کتابخانه rembg: این کتابخانه متن باز برای حذف بک گراند با استفاده از الگوریتم های غیر مبتنی بر OCR متن استفاده می شود ، همانطور که مشخص است در تصویر زیر بجای متن تنها شخص موجود در تصویر سگمنت شده است.





در تصویر بالا هم که متن دستنویس فارسی وجود دارد همراه با کجی! کتابخانه موجود نتوانسته آن را تحلیل و سگمنت کند و کل تصویر را خراب کرده است.

شماره	تاریخ	نام	لی	تعداد	قیمت کل
413	98/10/10	پیتزا مخلوط مرغ و گوشت	33500	1	33,500
		پیتزا استیک بوقلمون	35000	1	35,000
		سیب زمینی اسپیشیال	25000	1	25,000
		مالشعیر استوایی	4000	1	4,000
		مالشعیر کلاسیک	4000	3	12,000
		جمع کل :			109,500
		قابل پرداخت :			109,500
		ابتدای مطهری پ ۳۷۹			تلفن: ۸۸۱۰۰۰۰۸

شماره	تاریخ	نام	لی	تعداد	قیمت کل
413	98/10/10	پیتزا مخلوط مرغ و گوشت	33500	1	33,500
		پیتزا استیک بوقلمون	35000	1	35,000
		سیب زمینی اسپیشیال	25000	1	25,000
		مالشعیر استوایی	4000	1	4,000
		مالشعیر کلاسیک	4000	3	12,000
		جمع کل :			109,500
		قابل پرداخت :			109,500
		ابتدای مطهری پ ۳۷۹			تلفن: ۸۸۱۰۰۰۰۸

در بعضی از موارد مانند رسید بالا را خوب اسکن کرده است، در کل مدل رفتار نسبتاً رندومی دارد اما در مواردی که بکگراند یکنواختی داشته باشیم صرفنظر از زبان موجود در تصویر، عملکرد قابل مقایسه ای دارد!

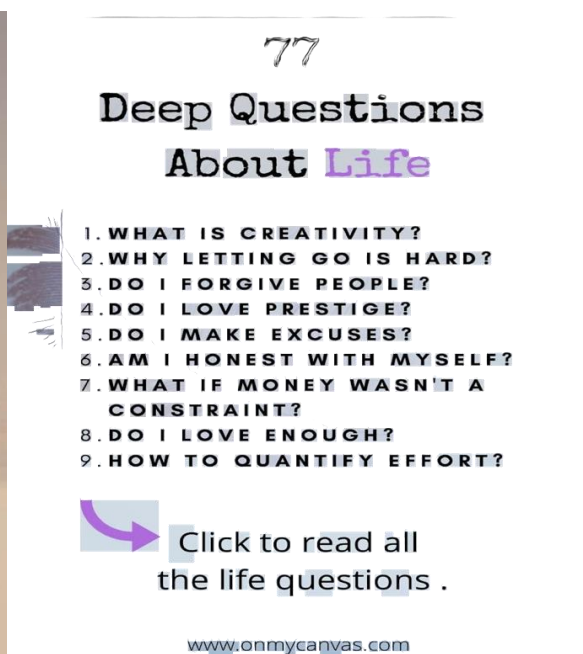
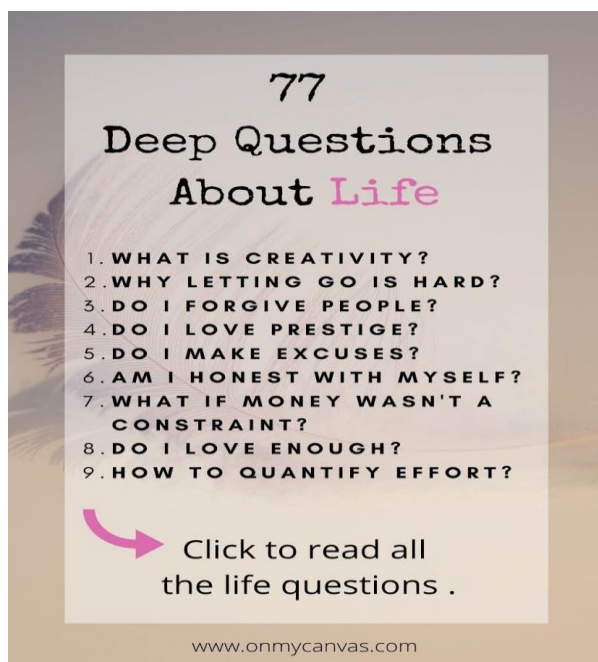
مزایا: استفاده از آن آسان و سریع است، همچنین با اعمال مراحل پیش پردازش و پس پردازش زاویه دید تصویر را تغییر می دهد.

معایب: حساسیت زیادی نسبت به کیفیت و شرایط تصویر دارد (مدل **robust** و **reliable** نیست، این موضوع در فروم ها اشاره شده است)، همچنین قابلیت خاصی برای فاین تیون کردن بر روی دیتاست مان را ندارد

مدل U2net: این مدل مبتنی بر یادگیری عمیق قابلیت بسیار بالایی در semantic segmentation دارد و بخوبی میتواند بکگراند را از تصاویر جدا کند به شرطی که بر روی دیتاستی به اندازه کافی بزرگ آموزش ببیند.

مزایا : قابلیت تحلیل معنایی تصاویر را دارد ، میتواند بخوبی برای تشخیص شی خاصی ، متن خاصی یا کاراکتر های خاصی در متن آموزش ببیند .

معایب: آموزش آنها منابع محاسباتی نسبتا زیادی می طلبد (البته می شود از سرویس های ابری آنلاین استفاده کرد) استقرار آنها هم نیاز به سیستم قدرتمندی دارد.(برای استقرار مدل روی گوشی یا سیستم های نهفته می شود با تکنیک هایی مثل quantization حجم و بار محاسباتی مدل نهایی را کاهش داد).



همینطور در شعر بالا و رسید تصویر زیر مشخص است ، مدل حساسیت زیادی نسبت به چرخش تصویر دارد که این موضوع را میتواند با اعمال داده افزایی (data augmentation) های مختلف و آموزش دوباره مدل رو آنها هندل کرد.

*استفاده ترکیبی از مدل ها(ensemble learning) : چون این مدل از OCR برای تشخیص متن استفاده نمیکند میتوان از ابتدا با استفاده از یک مدل قوی حوزه OCR مانند easy OCR استفاده کرد و با مکانیزم attention دقت مدل را برای سگمنت کردن بیشتر متوجه بخش هایی که دارای متن هستند جلب کرد.

شماره : 413		فهرست مواد		23:05:32 98/10/10	
ترک دانه					
35,500	1	35500	پیتزا مخلوط مرغ و گوشت		
35,000	1	35000	پیتزا استیک بوقلمون		
25,000	1	25000	سیب زمینی اسپیشیال		
4,000	1	4000	مالشعیر استوایی		
12,000	3	4000	مالشعیر کلاسیک		
109,500			جمع کل :		

قابل پرداخت : **109,500** تومان

ابتدای مطهری پ ۳۷۹ تلفن : ۸۸۱۰۰۰۰۸

شماره : 413		فهرست مواد		23:05:32 98/10/10	
ترک دانه					
35,500	1	35500	پیتزا مخلوط مرغ و گوشت		
35,000	1	35000	پیتزا استیک بوقلمون		
25,000	1	25000	سیب زمینی اسپیشیال		
4,000	1	4000	مالشعیر استوایی		
12,000	3	4000	مالشعیر کلاسیک		
109,500			جمع کل :		

قابل پرداخت : **109,500** تومان

ابتدای مطهری پ ۳۷۹ تلفن : ۸۸۱۰۰۰۰۸

الگوریتم **mean shift** : کاربرد اصلی این الگوریتم برای خوشه بندی است که برای سگمنتیشن تصاویر هم دقت بسیار مناسبی دارد
مزایا : الگوریتم ساده ای است که با تنظیم پارامتر های آن می توان نتایج خوبی برای تصاویر با کنتراست مناسب انتظار داشت

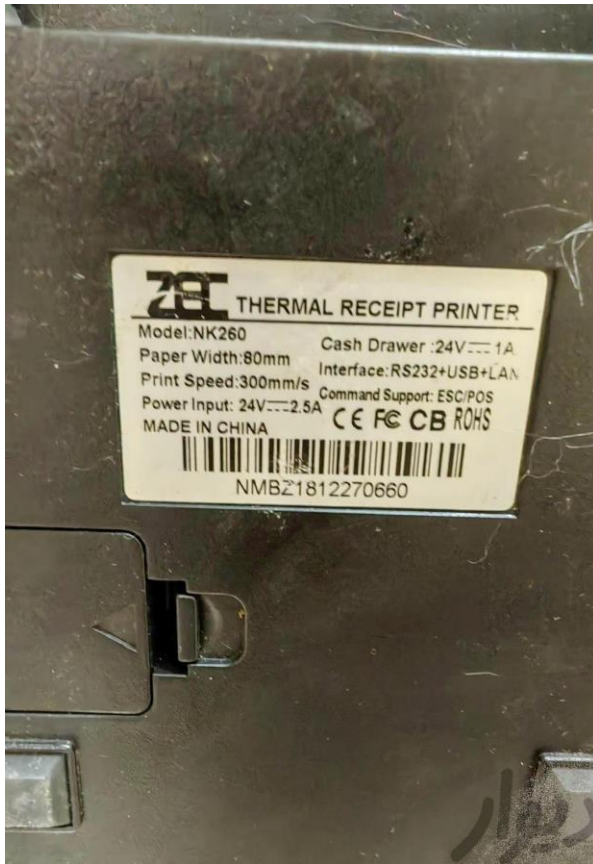


نقاط ضعف : برای تصاویر که یک دست نباشند (مثلا بخشی از برگه با سایه پوشیده شده باشد) یا کنتراست مناسبی نداشته باشند خوب عمل نمی کند و متن را حذف میکند.

Passargad Bank
Deposit Bill
From Date 23/07/2014 To Date 27/08/2014
Customer Name: [REDACTED]
Branch: [REDACTED]
Currency: IRR
Branch Code: 201
Teller: 2684
Date: 27/08/2014
Time: 10:25:37
Page: 1 Out Of 1

Row	Date	Time	Debit	Vouch. No.	Credit	Debit	Debit	Balance	Branch
1	23/07/14	10:00	-	307819	1,200,000,000	0	0	925,000	308
2	21/07/14	01:24	-	524112	2,300,310	0	0	1,200,310	308
3	02/08/14	11:37	-	908730	2,000,000	0	0	1,200,310	308
4	02/08/14	18:32	-	1004631	400,000	0	0	1,200,310	308
5	11/08/14	09:54	-	1035551	0	300,000	0	1,200,310	308
6	11/08/14	09:54	-	1035552	0	200,000	0	1,200,310	308
7	12/08/14	21:51	-	1064830	350,000	0	0	1,200,310	308
8	12/08/14	21:52	-	1067664	2,000,000	0	0	1,200,310	308
9	12/08/14	12:40	-	1088078	900,000	0	0	1,200,310	308
10	12/08/14	11:41	-	1101035	0	400,000	0	1,200,310	308
11	12/08/14	11:41	-	1124055	220,000	0	0	1,200,310	308
12	12/08/14	11:41	-	1181320	0	3,000,000	0	1,200,310	308
13	12/08/14	11:41	-	1183310	370,000	0	0	1,200,310	308
14	12/08/14	11:41	-	1183310	1,211,619,110	0	0	1,200,310	308
15	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
16	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
17	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
18	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
19	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
20	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
21	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
22	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
23	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
24	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
25	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
26	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
27	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
28	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
29	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
30	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
31	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
32	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
33	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
34	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
35	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
36	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
37	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
38	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
39	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
40	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
41	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
42	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
43	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
44	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
45	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
46	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
47	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
48	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
49	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
50	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
51	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
52	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
53	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
54	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
55	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
56	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
57	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
58	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
59	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
60	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
61	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
62	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
63	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
64	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
65	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
66	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
67	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
68	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
69	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
70	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
71	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
72	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
73	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
74	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
75	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
76	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
77	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
78	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
79	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
80	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
81	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
82	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
83	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
84	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
85	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
86	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
87	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
88	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
89	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
90	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
91	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
92	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
93	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
94	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
95	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
96	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
97	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
98	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
99	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308
100	12/08/14	11:41	-	1183310	1,200,000,000	0	0	1,200,310	308

Bank Passargad
Branch: Central
Code: 201



مدل Deep lab : این مدل مبتنی بر شبکه عصبی عمیق هست و قطعه بندی معنایی انجام می دهد و صرفا بر اساس بافت تصویر تصمیم گیری نمی کند و در این زمینه دقت بسیار بالایی دارد
مزایا : می تواند با استفاده از دیتاست های موجود OCR مدل را فاین تیون کرد.(اینکار نیازمند داشتن دیتاست و نسبتا زمانبر است)

نقاط ضعف : مدل پایه آن در حال حاضر قادر به تشخیص اسناد اسکن شده نیست و از تصاویر بخش هایی که اجسام (مطابق لیست تعریف شده اش) یا افراد باشند را سگمنت می کند.



در بقیه موارد تست شده هم که فقط برگه در کل تصویر باشد، هیچ چیز را سگمنت می کند.

آموزش آنها منابع محاسباتی نسبتا زیادی می طلبد (البته می شود از سرویس های ابری آنلاین استفاده کرد)
استقرار آنها هم نیاز به سیستم قدرتمندی دارد.(می شود با تکنیک هایی مثل quantization حجم و بار محاسباتی مدل نهایی را کاهش داد.

نکته نهایی : همچنین الگوریتم های خوشه بندی دیگر مانند k-means و dbscan نیز در این تست بررسی شدند که چون عملکرد خوبی نداشتند (و مورد انتظار هم بود) در این گزارش خروجی های آنها آورده نشده است.