



Assignment NO.9 Solutions

Digital Image Processing | Fall 1400 | Dr.Mohammadi

Teacher Assistant(s) : Fatemeh Anvari – Fatemeh Hajizadeh

Student name : **Amin Fathi**

Student id : **400722102**

Problem 1

R-CNN از آنجا که با ۲۰۰۰ پروپوزال در شبکه cnn کار می کند بنابراین برای هر تصویر باید ۲۰۰۰ پروپوزال استخراج شود و تحویل شبکه کانولوشنی شود که سرعت بسیار پایینی دارد Fast R-CNN نسخه سرعت یافته قبلی است و به جای دادن خروجی پروپوزال ها به شبکه کانولوشنی ، خود تصویر را به cnn می دهد و در نتیجه خبری از ۲۰۰۰ هزار پروپوزالی که سرعت را کند کرده بودند نیست که البته این دلیل دقت عملکرد این شبکه را پایین می آورد . در این مدل بعد از مرحله کانولوشن پروپوزال ها را تشخیص داده و به fc پاس می دهیم . دو الگوریتم R-CNN و Fast R-CNN برای پیدا کردن نواحی پروپوزال از الگوریتم selective search استفاده میکنند که الگوریتم زمانبری است بنابر این طراحان الگوریتم Faster R-CNN تصمیم گرفتند که این الگوریتم را از پروسه تشخیص object حذف کنند و به شبکه اجازه دهند که به طور خودکار نواحی پروپوزال را بیابد. در این الگوریتم مانند R-CNN تصویر به یک شبکه کانولوشنی داده میشود تا feature map ها استخراج شوند اما پس از استخراج feature map ها آنها را به selective search پاس نمیدهد و به جای اینکار یک شبکه کانولوشنی جداگانه طراحی کردند تا نواحی پروپوزال را پیشبینی کند و پس از reshape کردن خروجی ها آنها را به یک الگوریتم کلاسیفیکیشن پاس می دهند. Yolo برخلاف دو الگوریتم قبلی که دو مرحله ای بودند و فقط قسمتی از تصویر را در نظر می گرفتند است .

در واقع الگوریتم YOLO تصویر را به یک شبکه $S*S$ تبدیل میکند که در هر خانه از این شبکه تعداد m محدوده وجود دارد. برای هر محدوده الگوریتم probability و offset تعیین میکند و جعبه هایی که شانسی بیشتر از threshold مشخص شده داشته باشند به عنوان object تشخیص داده میشوند. الگوریتم YOLO چندین برابر از سایر الگوریتم ها سریع تر است ولی ضعف آن در کارکردن با اشیاء کوچک داخل تصویر است .

Ssd در واقع تصویر را به چند بلوک تقسیم می کند و چک میکند هر بلوک که آیا در خودش شی ای دارد یا نه که اگر نباشد حکم background دارد

Retina net یک مدل تشخیص اشیاء تک مرحله ای است که از loss ای به نام focal loss برای بررسی عدم تعادل در کلاس ها در حین training استفاده می کند

منابع :

[Object Detection Explained: R-CNN | by Chingis Oinar | Towards Data Science](#)

[R-CNN, Fast R-CNN, Faster R-CNN, YOLO — Object Detection Algorithms | by Rohith Gandhi | Towards Data Science](#)

[RetinaNet Model for object detection explanation - TowardsMachineLearning](#)

[How single-shot detector \(SSD\) works? | ArcGIS Developer](#)

[RetinaNet Explained | Papers With Code](#)

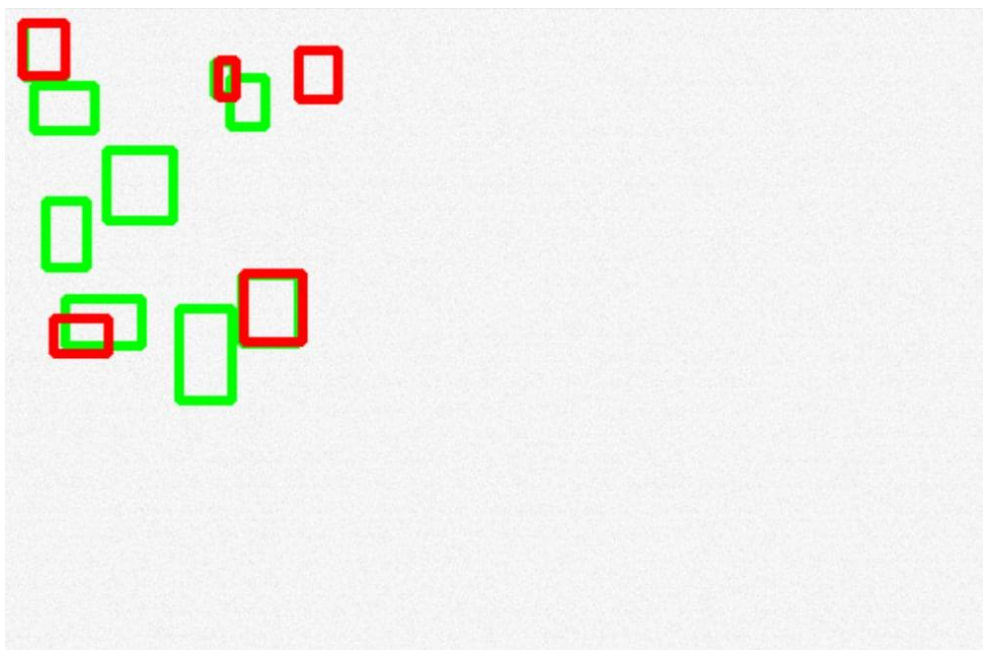
Problem 2

در این مقاله برای تولید object bounding box proposal ها از روش جدید به نام edges استفاده شده است. استفاده از edge ها مزایای محاسباتی زیادی را ارائه می دهد، زیرا ممکن است به طور موثر محاسبه شوند و نقشه های edge به دست آمده sparse هستند. در این مقاله، نحوه شناسایی مستقیم object proposal ها از نقشه های edge را بررسی می کنند. کار اصلی آن ها مشاهدات زیر است: تعداد خطوطی که به طور کامل توسط یک box محصور شده اند، نشان دهنده احتمال وجود یک شی در box است. اگر تمام پیکسل های edge متعلق به contour در داخل box قرار داشته باشند، می گویند یک contour به طور کامل توسط یک box محصور شده است. edge ها معمولاً با مرزهای شی مطابقت دارند و به همین دلیل جعبه هایی که مجموعه ای از لبه ها را قویاً محصور می کنند احتمالاً حاوی یک شی هستند. با این حال، برخی از edge هایی که در box مرزی یک شی قرار دارند، ممکن است بخشی از شی موجود نباشند. به طور خاص، پیکسل های لبه ای که متعلق به خطوطی هستند که مرزهای box را در بر می گیرند، احتمالاً با اشیاء یا ساختارهایی مطابقت دارند که در خارج از box قرار دارند.

بنابر نتایج این مقاله امتیاز دادن به یک box بر اساس تعداد خطوطی که به طور کامل در بر می گیرد، یک معیار proposal موثر ایجاد می کند. در مقابل، شمارش تعداد پیکسل های edge در box چندان موثر نیست. رویکرد این مقاله تا حدی شباهت هایی به معیارهای superpixels straddling دارد. با این حال، به جای اندازه گیری تعداد خطوط straddling، چنین خطوطی را در نظر نمی گیرند. از آنجایی که تعداد کادرهای محدود کننده ممکن در یک تصویر زیاد است، باید بتوان به طور موثر به نامزدها امتیاز داد. آن ها از Structured Edge detector استفاده می کنند تا نقشه edge اولیه را بدست آورند. برای کمک به محاسبات بعدی، پیکسل های edge همسایه با جهت گیری مشابه در کنار هم قرار می گیرند تا گروه ها را تشکیل دهند.

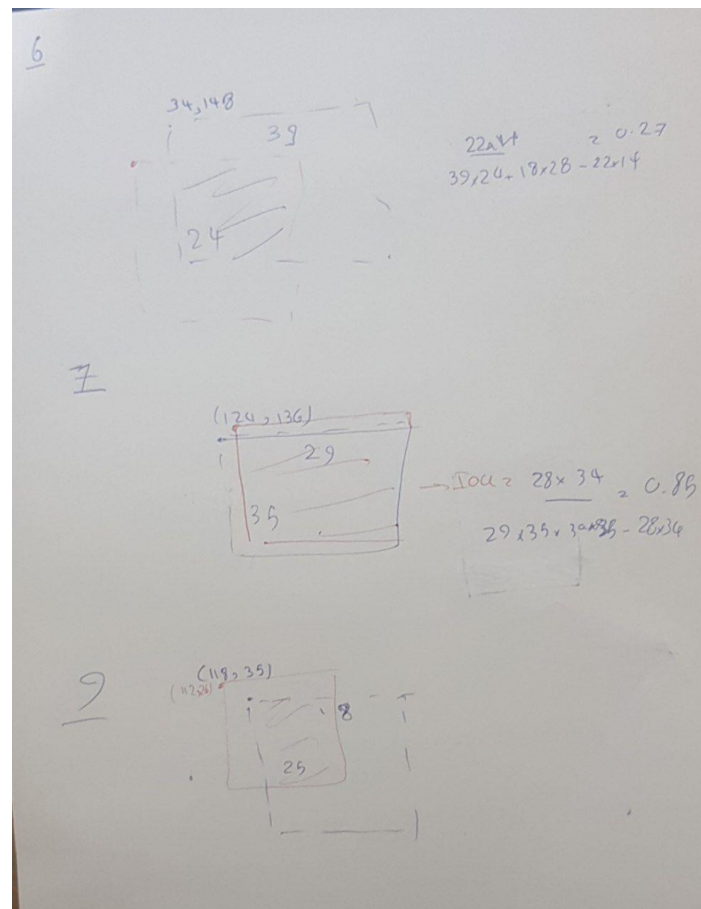
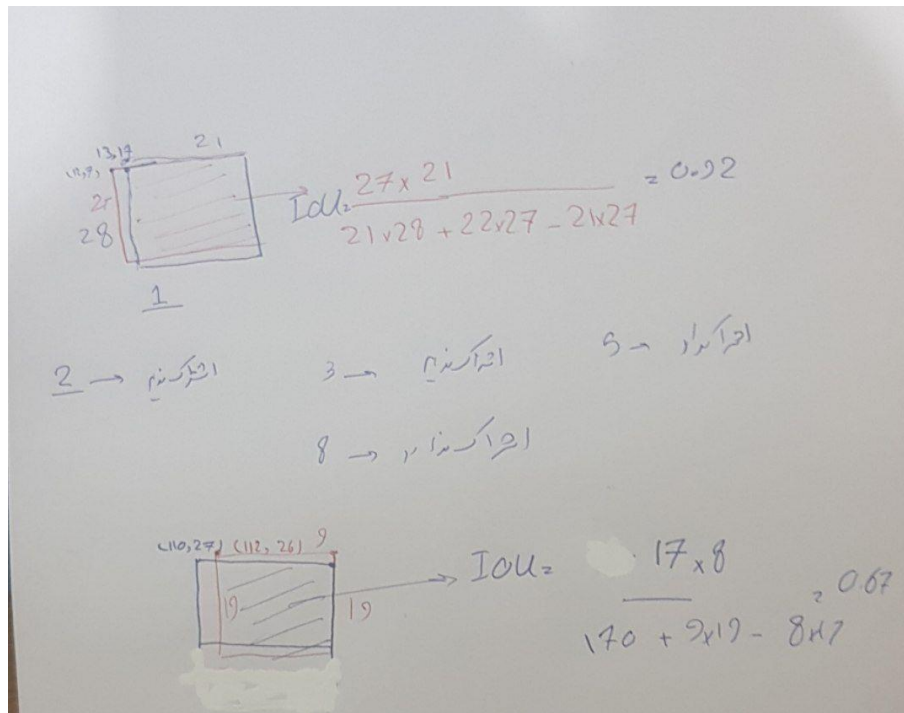
Problem 3

ابتدا دیتای داده شده را رسم میکنیم ، باکس های سبز ، باکس های پیش بینی هستند و باکس های قرمز باکس های
: garnd truth



همانطور که مشاهده میشود ۵ باکس قرمز وجود دارد ، ۹ باکس سبز هم در تصویر هست که ۳ باکس از آن ها با باکس های قرمز تطبیق حداکثری دارند و با کمی دقت قابل مشاهده هستند .

مقدار iou را برای سمپل هایی که داریم حساب میکنیم که برابر است با مساحت مشترک تقسیم بر مساحت کل .



به ترتیب ترشولد iou را برابر 0.25 0.5 0.75 قرار میدهیم و چنانچه iou نمونه ای از این مقدار بیشتر بود tp حساب شده و در غیر این صورت fp است .

سپس با توجه به فرمول زیر :

Definition of terms:

- True Positive (TP) — Correct detection made by the model.
- False Positive (FP) — Incorrect detection made by the detector.
- False Negative (FN) — A Ground-truth missed (not detected) by the object detector.

$$P = \frac{TP}{TP + FP} = \frac{TP}{\text{all detections}}$$
$$R = \frac{TP}{TP + FN} = \frac{TP}{\text{all ground-truths}}$$

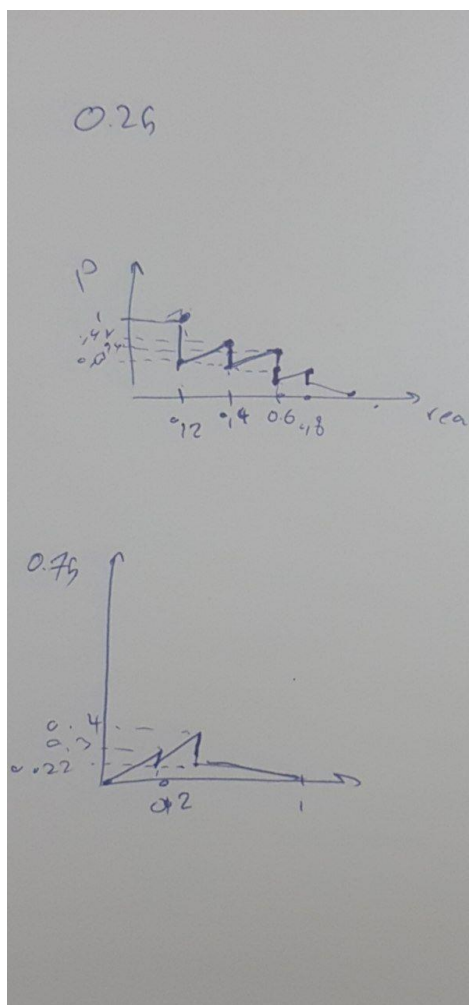
Equation 1: Precision and Recall

مقادیر precision و ریکال را برای هر نمونه به ترتیب از score بالا حساب می کنیم .

که در نتیجه محاسبات به شکل زیر میشود.

	y	w	h	score	IOU	0.25 P	R	0.5 P	R	0.75 P	R
13	7	21	28	0.84	0.67 TP	1	0.2 TP	1	0.2 FP	0	0
18	39	31	23	0.79	0 FP	0.5	0.2 FP	0.5	0.2 FP	0	0
55	72	34	36	0.89	0.92 TP	0.67	0.4 TP	0.67	0.4 TP	0.33	0.2
110	27	10	17	0.96	0 FP	0.5	0.4 FP	0.5	0.4 FP	0.25	0.2
24	98	21	34	0.47	0.85 TP	0.6	0.6 TP	0.6	0.6 TP	0.4	0.4
34	148	39	24	0.39	0.05 FP	0.5	0.6 FP	0.5	0.6 FP	0.33	0.4
124	136	29	35	0.74	0 FP	0.43	0.6 FP	0.43	0.6 FP	0.28	0.4
92	153	27	47	0.29	0.27 TP	0.5	0.8 FP	0.37	0.6 FP	0.25	0.4
118	35	18	25	0.62	0 FP	0.44	0.8 FP	0.33	0.6 FP	0.22	0.4

اگر که نمودار ها را بخواهیم رسم کنیم به شکل زیر می شود.



0.5

