



دانشگاه صنعتی امیر کبیر
(پلی تکنیک تهران)

درس پیتاپی کامپیوونر

استاد درس جناب آقای دکتر صفابخش

(تمرین سری ششم)

محسن عبادپور | ۴۰۰۱۳۱۰۸۰ | m.ebadpour@aut.ac.ir

نیمسال اول سال تحصیلی ۱۴۰۱-۱۴۰۲



فهرست پاسخ ها

مسئله اول: چگونگی استخراج ویژگی از نواحی مختلف توسط LBP ۳
مسئله دوم: امکان استفاده از استخراج ویژکی مبتنی بر نقاط کلیدی ۳
مسئله سوم: پیاده‌سازی ۴
مسئله سوم: استخراج ویژگی ۴
مسئله سوم: دسته‌بند ۴
مسئله سوم: نتایج / پیاده‌سازی سرکوب غیر حداکثرها (قسمت اضافی) ۵
مسئله سوم: نقاط ضعف و مشکلات ۶
مسئله چهارم: مدل رگرسیونی ۱۷
مسئله پنجم: سرکوب غیرحداکثرها ۱۱

مسئله اول: چگونگی استخراج ویژگی از نواحی مختلف توسط LBP

می‌دانیم که اساس ویژگی‌های استخراجی توسط LBP بر اساس تعداد پیکسل‌های موجود در تصویر می‌باشد و به ازای هر پیکسل یکبار استخراج ویژگی صورت می‌گیرد. از این رو و از جایی که نواحی استخراجی ابعاد و مساحت مختلفی دارند، بایستی تدبیری اندیشیده شود که طول بردار ویژگی از ابعاد ناحیه‌ی استخراجی مستقل شود.

همانطور که در کلاس مورد بررسی قرار گرفت، یک پیشنهاد مناسب برای این منظور استفاده از هیستوگرام حاصل از LBP‌های بدست آمده می‌باشد که در این صورت همواره طول بردار ویژگی ثابت خواهد بود. البته پیشنهاد‌های مختلف دیگری نظیر میانگین‌گیری یا ماکسیمم‌گیری نیز میتوان استفاده نمود اما اطلاعات حاصل از آن برای دسته‌بندی مفید نمی‌باشد. برای پیاده‌سازی از هیستوگرام LBP استفاده شده است. نکته حائز اهمیت نرمال‌سازی این هیستوگرام حاصل می‌باشد که بایستی صورت گیرد چرا که تعداد پیکسل‌های نواحی مختلف فرق داشته و در نتیجه رنج تغییرات هیستوگرام نیز متفاوت می‌باشد؛ نرمال‌سازی این مورد را حل می‌کند.

مسئله دوم: امکان استفاده از استخراج ویژگی مبتنی بر نقاط کلیدی

استفاده از استخراج ویژگی مبتنی بر نقاط کلیدی نظیر SIFT مناسب نمی‌باشد زیرا همانطور که در کلاس درس دیده شد، ویژگی‌های حاصل از اینان به تعداد نقاط کلیدی شناسایی شده وابسته بوده و برای هر ناحیه تعداد مختلفی نقاط کلیدی شناسایی می‌شود^(۱) که خود این زمانی بوده و سرعت کافی را ارائه نمی‌دهد^(۲) و همچنین هر یک از نقاط کلیدی نیز دارای یک بردار ویژگی مستقل خواهد بود؛ در نتیجه طول بردار ویژگی که حاصل از الحاق (concatenate) بردار‌های ویژگی نقاط کلیدی برای یک ناحیه می‌باشد ثابت نبوده^(۳) و روش‌های نظیر هیستوگرام یا میانگین‌گیری هم نمی‌تواند مفید واقع شود^(۴). همچنین ویژگی‌های استخراجی مبتنی بر نقاط کلیدی زمانی مناسب هستند که فرآیند تطبیق مدنظر باشد در صورتی که در مسئله‌ی تشخیص اشیا تطبیق نقاط کلیدی میسر نمی‌باشد^(۵)؛ با فرض محال اگر بخواهیم از تطبیق نقاط کلیدی هر یک از ROIs های آموزشی با هر یک از ROIs های آزمون استفاده نماییم، بصورت نمایی زمان پردازشی بالا^(۶) رفته و هیچ توجیهی نیز برای استفاده آن وجود نخواهد داشت. به نظر بnde مهمترین مشکل استفاده از استخراج ویژگی مبتنی بر نقاط کلیدی تفاوت در تعداد نقاط کلیدی در هر ناحیه و در نتیجه تفاوت در ابعاد بردار ویژگی حاصل برای هر ناحیه می‌باشد که ادعام و الحاق انان نیز بسیار چالشی می‌باشد.

مسئله سوم: پیاده‌سازی

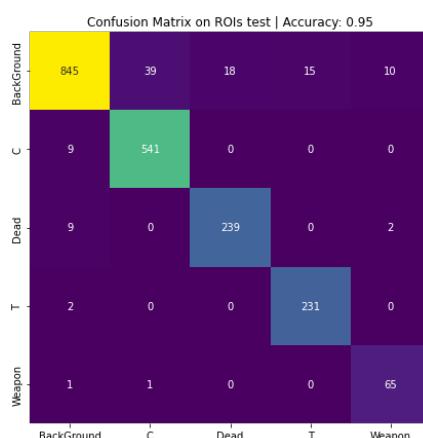
پیاده‌سازی های مورد نظر شامل استخراج ویژگی، آموزش دسته‌بند و همچنین آزمون تصاویر جدید انجام پذیرفته است.

مسئله سوم: استخراج ویژگی

برای استخراج ویژگی روش های مختلفی آزمایش شد که حاصل استفاده از ویژگی های Edge Density، LBP و HoG می‌بود. ابتدا هر یک به تنهایی با پارامترهای مختلف مورد آزمایش قرار گرفت و سپس ترکیب آنان نیز مورد ارزیابی قرار گرفت. بهترین نتیجه‌های که حاصل شد، حاصل الحق سه LBP با پارامتر های مختلف بود که به ترتیب برابر با (۳۲و۳)، (۱۶و۳) و (۵۴و۳) بود که اولی شعاع و دومی تعداد نقاط می‌باشد.تابع داده شده به گونه‌ای تغییر یافت که ورودی آن شامل تصویر و مختصات دو نقطه‌ی گوش ROI می‌باشد. استخراج سه ویژگی LBP اندکی زمانبر بوده ولی کیفیت آن طبق آزمایش های صورت گرفته بهتر می‌باشد لذا برای ارائه نتایج از آن استفاده شده است؛ اما اگر بخواهیم معیار سرعت را وارد مسئله کنیم، اولین بردار استخراجی می‌تواند کاملاً کافی و وافی باشد.

مسئله سوم: دسته‌بند

برای دسته‌بندی ROIs سه دسته‌بند مختلف مورد آزمایش قرار گرفت که شامل نزدیکترین همسایه، بردار های پشتیبان و درخت تصادفی مورد ارزیابی قرار گرفت. برای هر یک پارامتر های مختلف مورد آزمایش قرار گرفت؛ بهترین نتیجه‌های که حاصل شد مربوط به نزدیکترین همسایه با معیار سنحش فاصله کسینوسی بود که در زیر ماتریس درهم‌ریختگی حاصل از دسته‌بند آورده شده است:



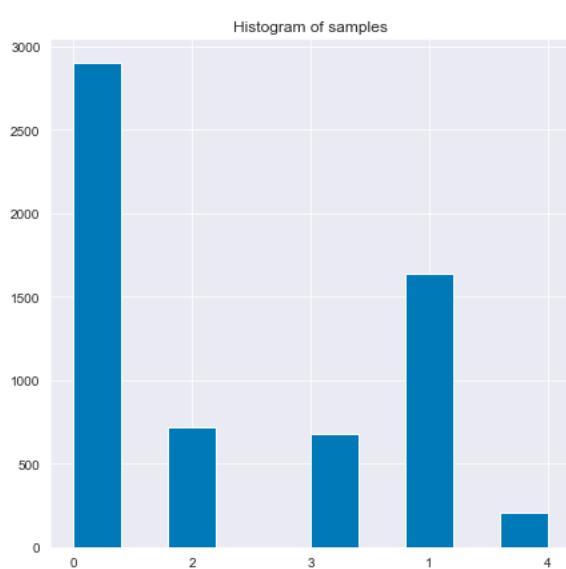
مسئله سوم: نتایج | پیاده سازی سرکوب غیر حداکثرها (قسمت اضافی)

بروی تمامی تصاویر داده شدهی آزمون فرآیند تشخیص اشیا صورت گرفته و در ادامه قابل مشاهده است. هر یک از نتایج شامل چهار تصویر اول ورودی مفروض می باشد؛ تصویر دوم شامل تمامی نواحی پیش بینی شدهی غیر پس زمینه می باشد. تصویر سوم مربوط به اعمال سرکوب غیر حداکثرها می باشد بطوری که این عمل بین نواحی پیش بینی شدهی هر کلاس بصورت جداگانه انجام پذیرفته است(طبق پیشنهاد مقاله)؛ برای مثال کل نواحی پیش بینی شده به عنوان اسلحه باهم در نظر گرفته شده است. تصویر چهارم نیز اعمال سرکوب غیر حداکثر بین کل نواحی پیش بینی شدهی غیر پیش زمینه اعمال شده است(برای بررسی و فعالیت بیشتر).



مسئله سوم: نقاط ضعف و مشکلات

اولین مشکل و نقطه ضعی در خروجی‌ها دیده می‌شود وقت نه چندان مناسب در اختصاص برچسب به نواحی در زمان آزمون نسبت به آموزش می‌باشد که دلیل آن دو مورد است: اولین دلیل نامناسب بودن ویژگی LBP و هیستوگرام حاصل از آن برای ویژگی‌های نواحی برای امر دسته بندی است؛ همانطور که در کلاس درس بحث شده است، LBP یک توصیفگر بافت بوده و به عنوان ویژگی تصویر در نظر گرفتنش کار صحیحی و اصولی نمی‌باشد. دومین دلیل نامتعادل (imbalance) بودن مجموعه داده‌های ROIs آموزشی (در متن سوال نخواسته شده است که این مشکل را حل و پیاده سازی کنیم و استخراج نواحی نامطلوب و نامرتبط برای تشخیص وجود اشیا می‌باشد که دسته بند را نیز در زمان آزمون با چالش مواجه می‌کند. نمودار فراوانی کلاس‌ها:



مشکل و نقطه ضعف بعدی تشخیص و شناسایی یک شی چندین بار با bounding box‌های مختلف می‌باشد که قابل انتظار بود چرا که نواحی پیشنهادی داخل هم بوده و شناسایی نیز در اکثر مواقع درست اتفاق افتاده است اما ادغام نتایج برای هر کلاس صورت نگرفته است بخاطر همین نتایج تو در تو حاصل شده است. مقاله برای رفع این مشکل سرکوب غیر حداقلی برای هر کلاس بصورت جداگانه را پیشنهاد داده است که بنده نیز این را اعمال نموده‌ام که در گزارش‌ها قابل ملاحظه است. مشکل و نقطه ضعف بعدی تشخیص با اختلاف bounding box‌های استخراجی در چندین مورد است. این مشکل در سوال بعدی مورد بررسی قرار می‌گیرد. مشکل و نقطه ضعف بعدی استخراج نواحی گاها بسیار نازک یا پهن و همچنین نواحی بسیار کوچک به جهت مساحت می‌باشد. برای حل این موضوع تلاش شده است با در نظر گرفتن یک شرط اینطور نواحی نادیده گرفته شود. این موضوع نشان میدهد با وجود دقت بالای دسته‌بند، استخراج نواحی نقص مواجه است.



V

Input image #3



All ROIs with detected classes



Non max suppression on ROIs based on predicted classes



Non max suppression on all ROIs that detected as non-background



A

Input image #4



All ROIs with detected classes



Non max suppression on ROIs based on predicted classes



Non max suppression on all ROIs that detected as non-background



Input image #5



All ROIs with detected classes



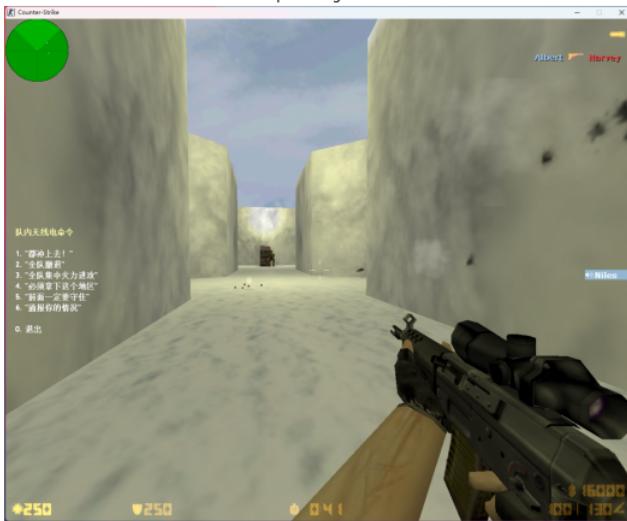
Non max suppression on ROIs based on predicted classes



Non max suppression on all ROIs that detected as non-background



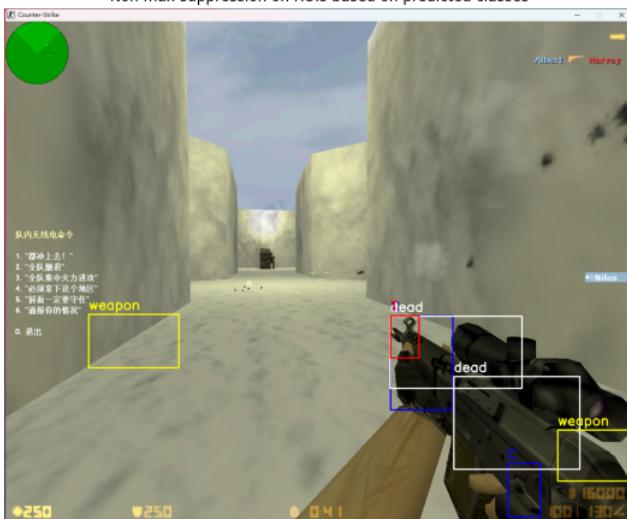
Input image #6



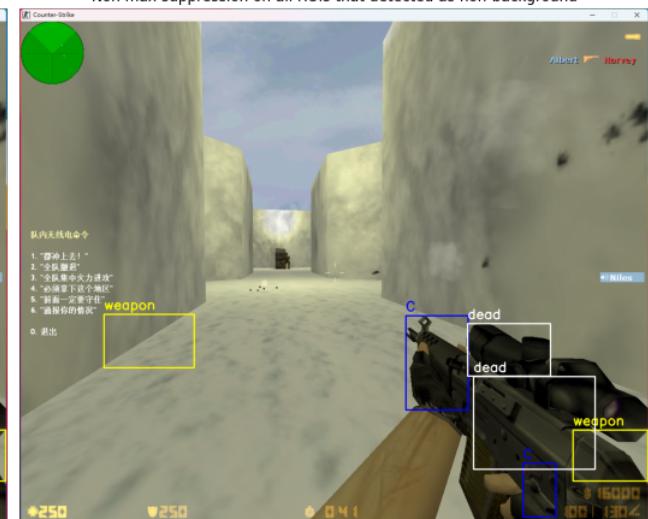
All ROIs with detected classes

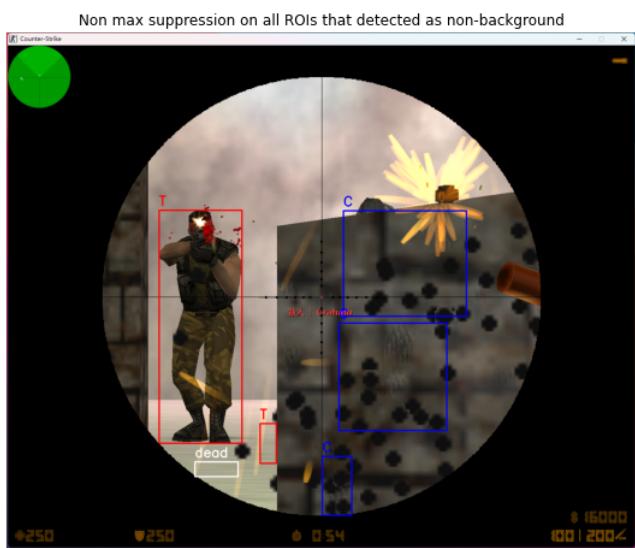
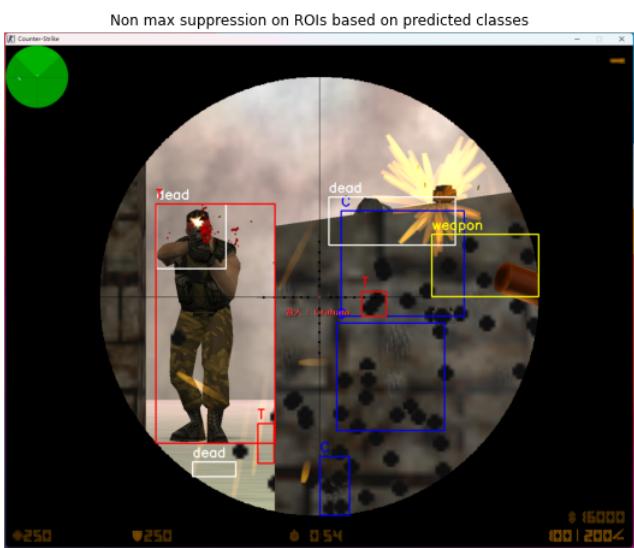
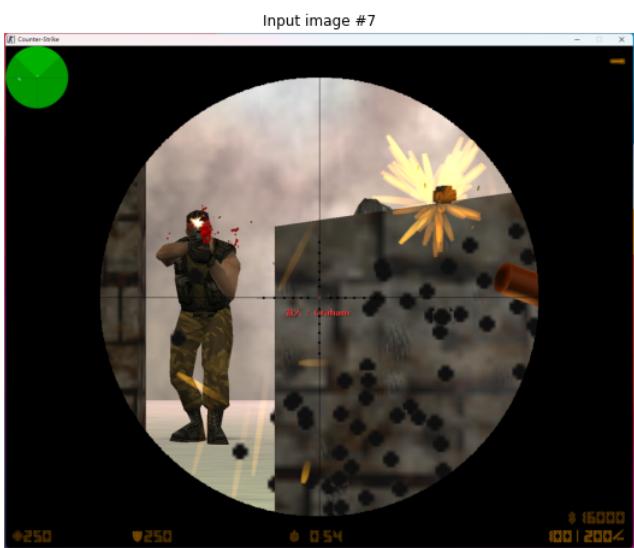


Non max suppression on ROIs based on predicted classes

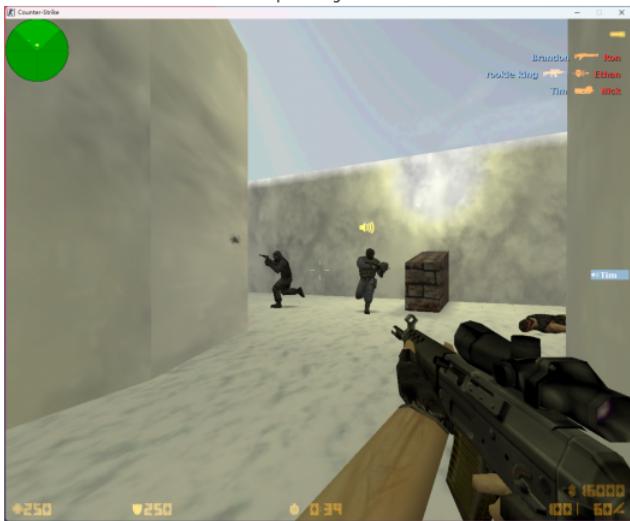


Non max suppression on all ROIs that detected as non-background

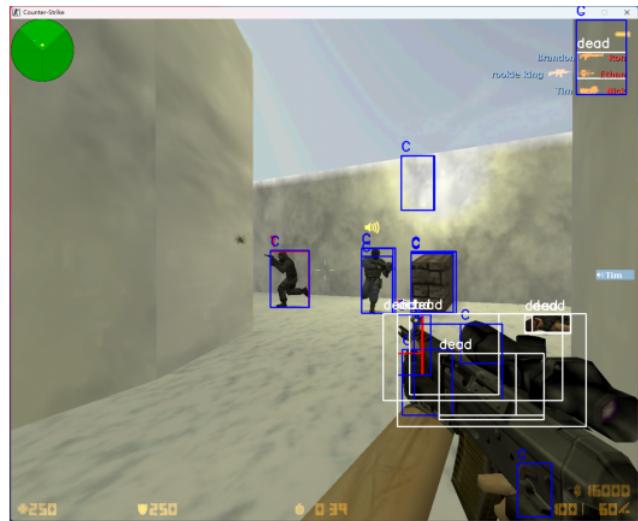




Input image #8



All ROIs with detected classes

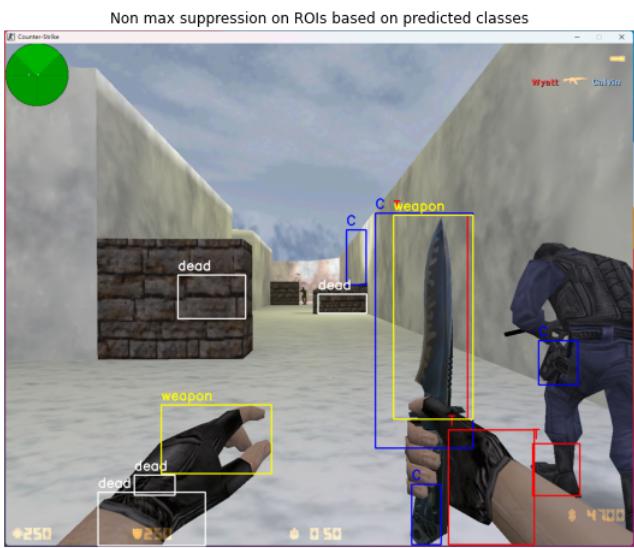


Non max suppression on ROIs based on predicted classes



Non max suppression on all ROIs that detected as non-background

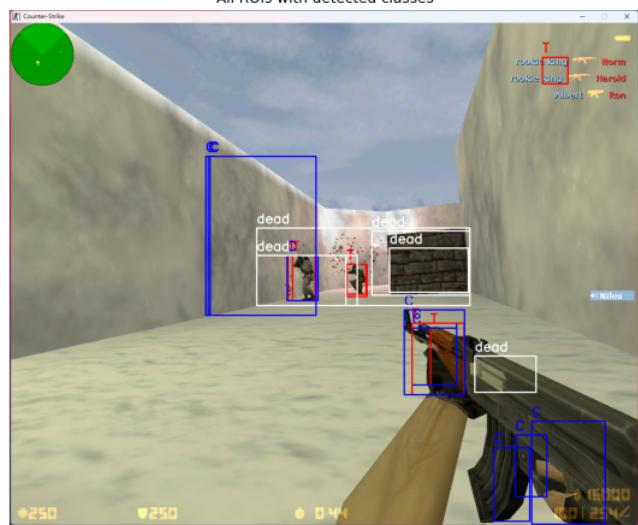




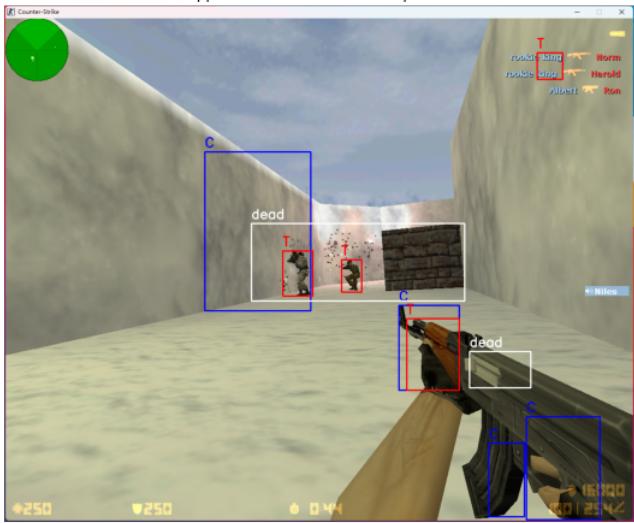
Input image #10



All ROIs with detected classes



Non max suppression on ROIs based on predicted classes



Non max suppression on all ROIs that detected as non-background



Input image #11



All ROIs with detected classes



Non max suppression on ROIs based on predicted classes



Non max suppression on all ROIs that detected as non-background



مسئله چهارم: مدل رگرسیونی

یکی از چالش هایی که شبکه ها و مدل های تشخیص اشیا با آن مواجه هستند، مکان یابی و اختصاص دقیق bounding box برای شی تشخیصی در تصویر ورودی می باشد. در الگوریتم R-CNN برای کاربرد مذکور که از اهمیت بالایی برخوردار است، از یک مدل رگرسیونی به ازای هر کلاس برای تخمین bounding box دقیق تخصیصی به شی استفاده می شود.

فرآیند این تخمین به این صورت اتفاق می افتد که ابتدا برچسب پیش بینی مشخص می شود؛ سپس بر اساس برچسب تخصیصی، مدل رگرسیون متناظر با آن کلاس بردار ویژگی حاصل از شبکه CNN را ورودی گرفته و اقدام به تخمین bounding box جدید و دقیق می کند. در بحث تشخیص اشیا دقت bounding box اختصاص داده شده از جهات مختلفی اهمیت و کاربرد دارد مثلا در بحث بینایی بازو های مکانیکی برای تشخیص محل قطعات، اگر محل شی و bounding box تخصیصی دقت کافی را نداشته باشد، بازو ممکن است نتواند قطعه را بردارد یا آن را زمین بندازد.

بنده این مورد را در چندین حالت و شرایط(تخمین با یک مدل، تخمین با چندین مدل، تخمین هر پارامتر با یک مدل و...) آزمایش کردم که نتیجه همگی آنان بسیار نامطلوب بوده و حتی امکان استفاده نیز وجود نداشت؛ برای مثال در موارد متعددی bounding box تخمینی توسط مدل رگرسیونی از ابعاد تصویر بیرون می زد) در زیر مقایسه مقدار مورد انتظار و تخمینی یکی از پارامتر های bounding box (offset) آورده شده است. قابل ملاحظه است که نتایج حتی نزدیک نیز نمی باشد. همانطور که قبل تر مورد اشاره قرار گرفت، مهمترین دلیل این مشکل، استفاده از هیستوگرام LBP که یک توصیفگر بافت می باشد، به جای بردار ویژگی است. به عنوان مثال هیستوگرام LBP اطلاعات همسایگی را در بردار نهایی ویژگی لحظ نمیکند لذا نمیتوان انتظار داشت بتوان تخصیص bounding box که بسیار مرتبط با داده های محلی و همسایگی پیکسل ها می باشد را بدستی انجام دهد.

[-25.41146018 -295.]

[-19.90722602 70.]

[-150.35889028 91.]

[173.0757955 350.]

[175.03848608 539.]

مسئله پنجم: سرکوب غیر حداکثرها

یکی از چالش‌های مشکلاتی که شبکه‌های تشخیص اشیا با آن مواجه هستند، اختصاص چندین bounding box با ابعاد مختلف برای یک شی در مرحله تست می‌باشد که می‌تواند بصورت تودرتو حاصل شود. در زمان تست ما حقیقت مبنای bounding box های یکتا را نداریم و از طرفی طبق نواحی پیشنهادی ممکن است چندین و چند ناحیه از یک شی پیشنهاد شده و همگی نیز به کلاس درست پیش بینی شوند که این باعث می‌شود ناحیه‌های یک شی یکتا چندین بار به عنوان آن شی شمارش و برای آن bounding box های متعدد مختلف خروجی داده شود که کاملاً اشتباه است. در مقاله R-CNN برای حل این چالش از سرکوب غیرحداکثری برای نواحی پیش بینی شده استفاده می‌شود. سرکوب غیر حداکثرها در مقاله مذکور بصورت مستقل بین کلاس‌ها اعمال می‌شود بطوریکه بین نواحی تشخیص یافته برای هر برچسب عملیات سرکوب غیر حداکثر صورت گرفته و ناحیه‌ی جامع‌تر در نهایت حاصل می‌شود. توضیح مربوط به این قسمت از مقاله در زیر آورده شده است. لازم به ذکر است که در مقاله مذکور حد آستانه‌ی اعمالی به عنوان یک پارامتر یاد گرفته می‌شود.

2.2. Test-time detection

At test time, we run selective search on the test image to extract around 2000 region proposals (we use selective search's "fast mode" in all experiments). We warp each proposal and forward propagate it through the CNN in order to compute features. Then, for each class, we score each extracted feature vector using the SVM trained for that class. Given all scored regions in an image, we apply a greedy non-maximum suppression (for each class independently) that rejects a region if it has an intersection-over-union (IoU) overlap with a higher scoring selected region larger than a learned threshold.

بنده این قسمت را به عنوان امتیاز و فعالیت اضافی پیاده سازی کرده ام که در نتایج قابل مشاهده است.