به نام خدا

دانشگاه صنعتی امیرکبیر

دانشکده مهندسی کامپیوتر

استاد درس: دکتر صفابخش

پاییز ۱۴۰۱

درس بینائی کامپیوتر

تمرین ششم

هدف: آشنایی با مسئله تشخیص اشیا و استفاده از ویژگی‌های دودویی محلی (LBP) برای حل آن

کد: در پیاده سازی می توانید از زبان های پایتون، متلب یا سی پلاس پلاس استفاده کنید. همچنین در تمامی موارد می توانید از کتابخانه اپن سی وی و scikit image استفاده کنید مگر اینکه صراحتا خلاف ان در صورت سوال ذکر شده باشد .

گزارش: توجه کنید ملاک اصلی برای ارزیابی گزارش تمرین می باشد. برای این منظور گزارش را در قالب PDFتهیه کنید و برای هر سوال، تصاویر ورودی، خروجی و توضیحات مربوط به ان را ذکر کنید. همچنین اگر فرض اضافه ای در نظر می گیرید حتما در گزارش به ان اشاره کنید .

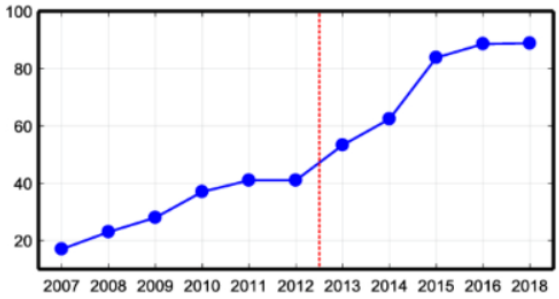
تذکر: مطابق قوانین دانشگاه هر گونه کپی برداری و اشتراک کار دانشجویان غیرمجاز بوده و شدیدا برخوردخواهد شد. استفاده از کدها و توضیحات اینترنت به منظور یادگیری بلامانع است، اما کپی کردن غیرمجازاست .

راهنمایی: در صورت نیاز سوالات خود را می توانید در گروه مربوط به درس در تلگرام یا با ایمیل زیر مطرح کنید.

E-mail: cv.ceit.aut@gmail.com

ارسال: فایل های کد و گزارش را در قالب یک فایل فشرده با فرمت studentID_HW06.zip تا تاریخ بیست دی‌ماه ارسال نمایید.
تاخیر مجاز: در طول ترم، مجموعا مجاز به حداکثر ده روز تاخیر برای ارسال تمرینات هستید(بدون کسر نمره). این تاخیر را میتوانید برحسب نیاز بین تمرینات مختلف تقسیم کنید، اما مجموع تاخیرات تمام تمرینات شما نباید بیشتر از ده روز شود. پس از استفاده از این تاخیر مجاز، هر روز تاخیر باعث کسر ده درصد نمره خواهد شد .

تشخیص اشیا[1] از مهم‌ترین مسائل علم بینایی ماشین، علیرغم سابقه طولانی، همچنان بخش زیادی از تلاش‌های محققین در این حوزه را به خود اختصاص داده است[2]. اگرچه با ظهور شبکه‌های کانولوشنی عمیق، الگوریتم‌های کلاسیک پردازش تصویر، دست‌کم از نظر دقت، توان رقابتی خود را از دست داده‌اند (شکل ۱)؛ اما همانطور



شکل 1- نمودار معیار ارزیابی mAP@50 برای مدلهای پیشنهادی تشخیص اشیا در طول زمان. خط عمودی قرمز رنگ نمایانگر ظهور شبکه‌های کانولوشنی عمیق است.

[1] Object detection

[2] Liu, Li, et al. "Deep learning for generic object detection: A survey." International journal of computer vision 128.2 (2020): 261-318.

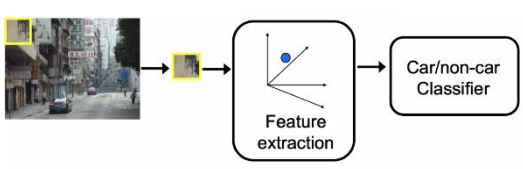
که در ادامه خواهیم دید، الگوریتم‌های مبتنی بر یادگیری عمیق نیز، علیرغم تلاش‌هایی در خلاف این جهت[3]، همچنان تا حد زیادی به ماژول‌های دست ساز طراحی شده توسط محققین وابسته هستند. بنابراین می‌توان نقش پررنگ مهندسین بینایی ماشین را در حل این مسئله به سادگی مشاهده کرد و از این منظر، الگوریتم‌های تشخیص اشیا، شباهت زیادی با روش‌های کلاسیک دارند. به همین جهت، حل یک مسئله تشخیص اشیا با روش‌های کلاسیک، که هدف از این تمرین است، می‌تواند شروع مناسبی برای ورود به روش‌های یادگیری عمیق در بینایی ماشین باشد.

بخش اول) آشنایی با رویکرد دو مرحله‌ای تشخیص اشیا

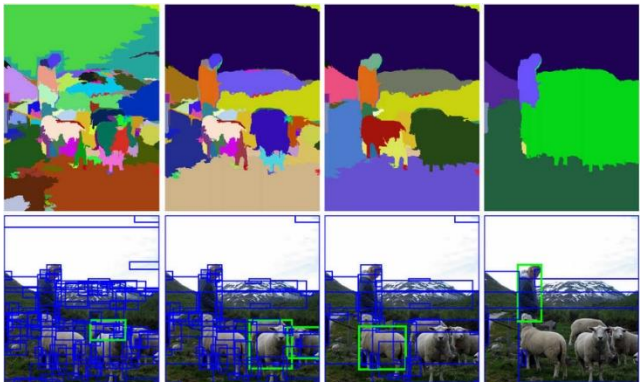
رویکرد دو مرحله‌ای تشخیص اشیا، از قدیمی‌ترین رویکردهای پیشنهادی برای حل مسئله شناسایی اشیا، چه در روش‌های کلاسیک و چه در روش‌های مبتنی بر یادگیری عمیق است. این رویکرد از گام های زیر تشکیل شده‌است.

گام اول) پیشنهاد نواحی

نخستین راهی که برای تشخیص اشیا موجود در یک تصویر به ذهن می‌رسد، قراردادن یک پنجره متحرک در تمام محل‌های ممکن در تصویر و آموزش یک دسته‌بند برای دسته‌بندی این پنجره‌ها است(شکل ۳). وجود اشیا با مقیاس‌های مختلف در تصویر، این روش را برای مسائل واقعی بسیار پرهزینه می‌کند. بنابراین به نظر می‌رسد نیاز به روش هوشمندانه‌ای برای پیشنهاد تعداد کمتری نواحی که احتمال وجود اشیا در آن بیشتر است، داریم[4]. الگوریتم



شکل 3-روش پنجره متحرک برای تشخیص اشیا



شکل 2-الگوریتم جستجوی انتخابی برای پیشنهاد ناحیه

[3] Carion, Nicolas, et al. "End-to-end object detection with transformers." European conference on computer vision. Springer, Cham, 2020.

[4] در ادبیات بینایی کامپیوتر به این نواحی، Region of Interest یا به اختصار ROI می‌گویند.

شکل 4- فرآیند تخصیص برچسب مبنا به نواحی پیشنهادی- چپ: نواحی مبنا و برچسب آن‌ها که توسط انسان مشخص می‌شوند. وسط: نواحی پیشنهادی الگوریتم جستجوی انتخابی. راست: نواحی انتخاب شده و برچسب آن‌ها که به عنوان داده آموزشی به مرحله بعد می‌روند. به سایر نواحی برچسب background اختصاص داده می‌شود(این نواحی در شکل راست برای جلوگیری از شلوغی نمایش داده نشده‌اند).

جستجوی انتخابی[5] را می‌توان اولین روش عملی برای پیشنهاد هوشمندانه نواحی دانست. این الگوریتم با خوشه‌بندی سلسه مراتبی تصویر ورودی، به تدریج نواحی با مقیاس‌های مختلف پیشنهاد می‌دهد(شکل ۳). تعداد نواحی پیشنهادی این الگوریتم معمولا به ندرت از ۱۰ هزار ناحیه فراتر می‌رود، که بسیار کمتر از روش پنجره متحرک خواهد بود.

گام دوم) تخصیص برچسب مبنا[6] به نواحی پیشنهادی (فقط برای مرحله آموزش)

پس از استخراج نواحی پیشنهادی توسط الگوریتم جستجوی انتخابی، باید برچسب مبنای آن‌ها نیز تعیین شود؛ زیرا این نواحی و برچسب آن‌ها در ادامه به عنوان داده آموزشی به یک شبکه دسته‌بند داده خواهد شد. به طور خلاصه نواحی پیشنهادی که به نواحی مبنا(شکل ۴-چپ) شبیه هستند، برچسب این نواحی را به خود می‌گیرند(شکل ۴-راست). سایر نواحی به عنوان background درنظر گرفته می‌شوند. بنابراین اگه مسئله تشخیص اشیا به عنوان مثال ۴ کلاسه باشد؛ دسته‌بند همواره یک کلاس اضافه background نیز خواهد داشت و بنابراین دیتاست ۵ کلاسه خواهد بود. در پایان ذکر این نکته ضروری است که از آنجا که تعداد داده‌های background در دیتاست آموزشی بسیار بیشتر از سایر کلاس‌ها است، معمولا از یک روش نمونه‌برداری برای جلوگیری از بایاس در دسته‌بند استفاده می‌شود[7].

گام سوم) استخراج ویژگی

پس از ساخت دیتاست، می‌توان نواحی پیشنهادی را مستقیما برای آموزش به یک دسته‌بند (مانند SVM) داد. با این حال، بهتر آن است که دسته‌بند را بر روی ویژگی‌های استخراج شده از این نواحی آموزش دهیم تا در برابر تغییرات روشنایی، حالت و ... مقاوم‌تر باشد. برای این کار می‌توان از هر روش دلخواهی برای استخراج ویژگی، از روش‌های کلاسیک تا شبکه‌های کانولوشنی عمیق استفاده کرد. تنها نکته مهم این است که طول بردار ویژگی استخراج شده باید صرفنظر از اندازه ناحیه ثابت باشد.

---

[5] Uijlings, Jasper RR, et al. "Selective search for object recognition." International journal of computer vision 104.2 (2013): 154-171.

[6] Ground truth

[7] برای اطلاعات بیشتر این عبارت را جستجو کنید: Sampling Methods for Imbalanced Classification

گام چهارم) آموزش دسته‌بند

درنهایت، به سادگی می‌توان با جفت ویژگی-برچسب‌های استخراج شده، یک دسته‌بند دلخواه آموزش داد. در این گام نیز برای انتخاب دسته‌بند آزادی عمل وجود دارد.

گام پنجم) استفاده از الگوریتم در مرحله تست

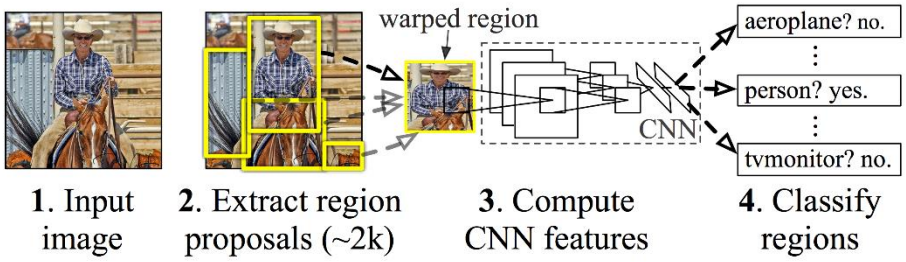
برای استفاده از الگوریتمی که تا این مرحله توسعه دادیم، پس از آموزش دسته‌بند، کافی است گام‌های گذشته را بر روی تصویر ورودی جدید اجرا کنید. توجه کنید که برای تصویر جدید در مرحله تست دسترسی به نواحی مبنا نداریم. پس از استخراج نواحی پیشنهادی(گام اول) و استخراج ویژگی از این نواحی(گام سوم)، همه آن‌ها را به وسیله دسته‌بندِ آموزش دیده، برچسب می‌زنیم. مطابق انتظار، دسته‌بند احتمالا بیشتر این نواحی را به عنوان background پیش‌بینی خواهد کرد که نیازی به نمایش آن‌ها وجود ندارد. سایر نواحی به همراه برچسب آن‌ها بر روی تصویر اصلی، خروجی الگوریتم تشخیص اشیا هستند.

پنج گامی که در بالا شرح داده شد، گام‌های اصلی الگوریتم مطرح RCNN[۸] هستند. این الگوریتم اگرچه به عنوان یکی اولین روش‌های مبتنی بر یادگیری عمیق شناخته می‌شود؛ اما همانطور که در شکل ۵ مشاهده می‌کنید، تنها در مرحله استخراج ویژگی از شبکه‌های عصبی عمیق استفاده می‌کند. بنابراین به راحتی می‌توان آن را با روش‌های کلاسیک جایگزین کرد. در ادامه با جزئیات پیاده سازی این پنج گام آشنا خواهیم شد.

بخش دوم) پیاده سازی یک الگوریتم تشخیص اشیا با ویژگی‌های LBP

دیتاست cstrike detection حاوی حدود ۲۰۰ عکس از بازی محبوب Counter-Strike به همراه برچسب‌های تشخیص اشیا مربوط به هر عکس با فرمت استاندارد YOLO (پیوست یک) است. این دیتاست به پیوست در اختیار شما قرار می‌گیرد.

هدف شما، پیاده‌سازی یک الگوریتم تشخیص اشیا با ویژگی‌های LBP برای حل مسئله تشخیص اشیا در این دیتاست است. پیاده‌سازی کامل این الگوریتم بسیار وقت‌گیرتر از حجم درنظر گرفته شده برای یک تمرین این درس است؛ بنابراین بخش زیادی از کد، شامل کدهای مربوط به خواندن داده‌ها، تهیه دیتاست و ... به صورت آماده در اختیار شما قرار می‌گیرد و شما فقط باید قسمت‌های مربوط به استخراج ویژگی از نواحی و آموزش دسته‌بند را انجام دهید. در هر حال آشنایی با نحوه عملکرد هر قسمت از کد، به درک شما از این مسئله کمک خواهد کرد. بنابراین پیشنهاد می‌شود کدها را به دقت بررسی کنید.



شکل 5-فرآیند کلی الگوریتم RCNN

فرآیند کلی الگوریتم مشابه آنچه در قسمت قبل توضیح داده شد، خواهد بود. با ذکر این نکات که:

- در گام دوم، برای تخصیص برچسب مبنای نواحی پیشنهادی، از معیار IoU(پیوست ۱) استفاده می‌شود. بدین ترتیب که نواحی پیشنهادی که IoU آن‌ها با یک ناحیه مبنا بیش از ۰٫۷ باشد، برچسب آن ناحیه را خواهد گرفت. نواحی پیشنهادی که IoU آن‌ها با هیچ ناحیه مبنایی بیش از ۰٫۳ نباشد، برچسب background خواهند داشت. سایر نواحی حذف شده و در آموزش شرکت داده نمی‌شوند.
- در گام چهارم-استخراج ویژگی- حتما باید ویژگی LBP جزء ویژگی‌های استخراج شده از نواحی پیشنهادی باشند. استفاده از سایر ویژگی‌ها مانند HoG در کنار LBP اختیاری می‌باشد.

با توجه به توضیحات فوق:

الف) به نظر شما چگونه می‌توان از ویژگی‌های LBP برای استخراج ویژگی از نواحی پیشنهادی استفاده کرد؟ چالش اصلی در اینجا این است که طول بردار ویژگی نواحی، صرفنظر از اندازه آن‌ها باید یکسان باشد. همچنین استخراج ویژگی باید تا حد ممکن با سرعت بالایی صورت بگیرد.

ب) آیا استفاده از روش‌های استخراج ویژگی مبتنی بر نقاط کلیدی(مانند SIFT) برای استخراج ویژگی از نواحی پیشنهادی، ممکن/مناسب است؟

ج) کدهای ارائه شده را با توجه به توضیحات کامل کنید، الگوریتم دسته‌بند را آموزش دهید و نتایج آن را روی تصاویر مجموعه تست ارائه دهید. مهم‌ترین نقطه‌ضعفی که در خروجی‌ها دیده می‌شود چیست؟

د-امتیازی) الگوریتم RCNN در کنار دسته‌بند، دارای یک مدل رگرسیون به ازای هر کلاس از دیتاست است. اهمیت و کاربرد این مدل‌های رگرسیون چیست؟ سازوکار مشابهی به الگوریتم خود اضافه کنید. می‌توانید به جای چند مدل، از یک مدل رگرسیون استفاده کنید.

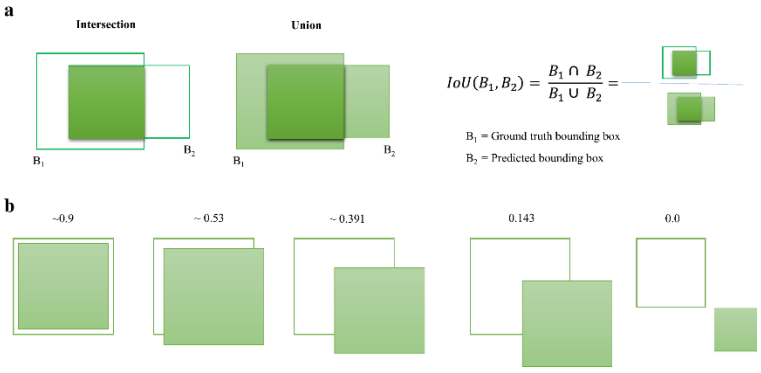
ه) استفاده الگوریتم سرکوب غیرحداکثری[9] که در درس با آن آشنا شدید، برای کدام گام از الگوریتم توضیح داده شده لازم است؟ (نیازی به پیاده‌سازی این بخش نیست)

[9] non maximum suppression

پیوست ۱

الف) معیار شباهت‌سنجی IoU

اندیس ژاکار که در ادبیات بینایی کامپیوتر بیشتر به عنوان [نسبت] اشتراک بر اجتماع[10] شناخته می‌شود، برای سنجش میزان شباهت دو کادر محصورکننده استفاده می‌شود.



شکل 6- معیار شباهت سنجی IoU

ب) فرمت استاندارد YOLO برای دیتاست‌های تشخیص اشیا

شبکه YOLO[11] احتمالا معروف‌ترین شبکه تشخیص اشیا برای اهداف صنعتی و تحقیقات آکادمیک خارج از حوزه بینایی ماشین است. بنابراین تعجب برانگیز نیست که فرمت مورد انتظار این شبکه برای آموزش آن، به استانداردی برای ذخیره دیتاست‌های تشخیص اشیا تبدیل شده‌است. این فرمت به ازای هر تصویر در دیتاست آموزش، دارای یک فایل متنی با نام یکسان و با ساختاری مشابه زیر است:

```
0 0.180404 0.596866 0.141524 0.261159
0 0.508554 0.542735 0.029549 0.101614
2 0.472395 0.528965 0.019440 0.055081
1 0.479782 0.600190 0.077760 0.039886
3 0.945179 0.653846 0.109642 0.027540
3 0.488725 0.633903 0.059876 0.016144
```

Class id x-center y-center width height

[10] intersection over union

[11] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

شکل 7-یک نمونه از فرمت استاندارد YOLO برای دیتاست‌های تشخیص اشیا

هر ردیف نماینده یک کادرمحصورکننده است که به وسیله عناصر فوق تعریف می‌شود. لازم به ذکر است همه اعداد نسبت به ابعاد تصویر نرمال شده‌اند. برای درک بهتر به تصویر ۷ توجه کنید. به محورهای مختصات این شکل و تفاوت آن با استاندارد آرایه‌های کتابخانه NumPy توجه کنید.