باسمه تعالی

دانشگاه صنعتی شریف

دانشکده مهندسی برق

دکتر فاطمی‌زاده -یادگیری عمیق
سوال امتیازی تیوری تمرین دوم

۱ سوال ۱

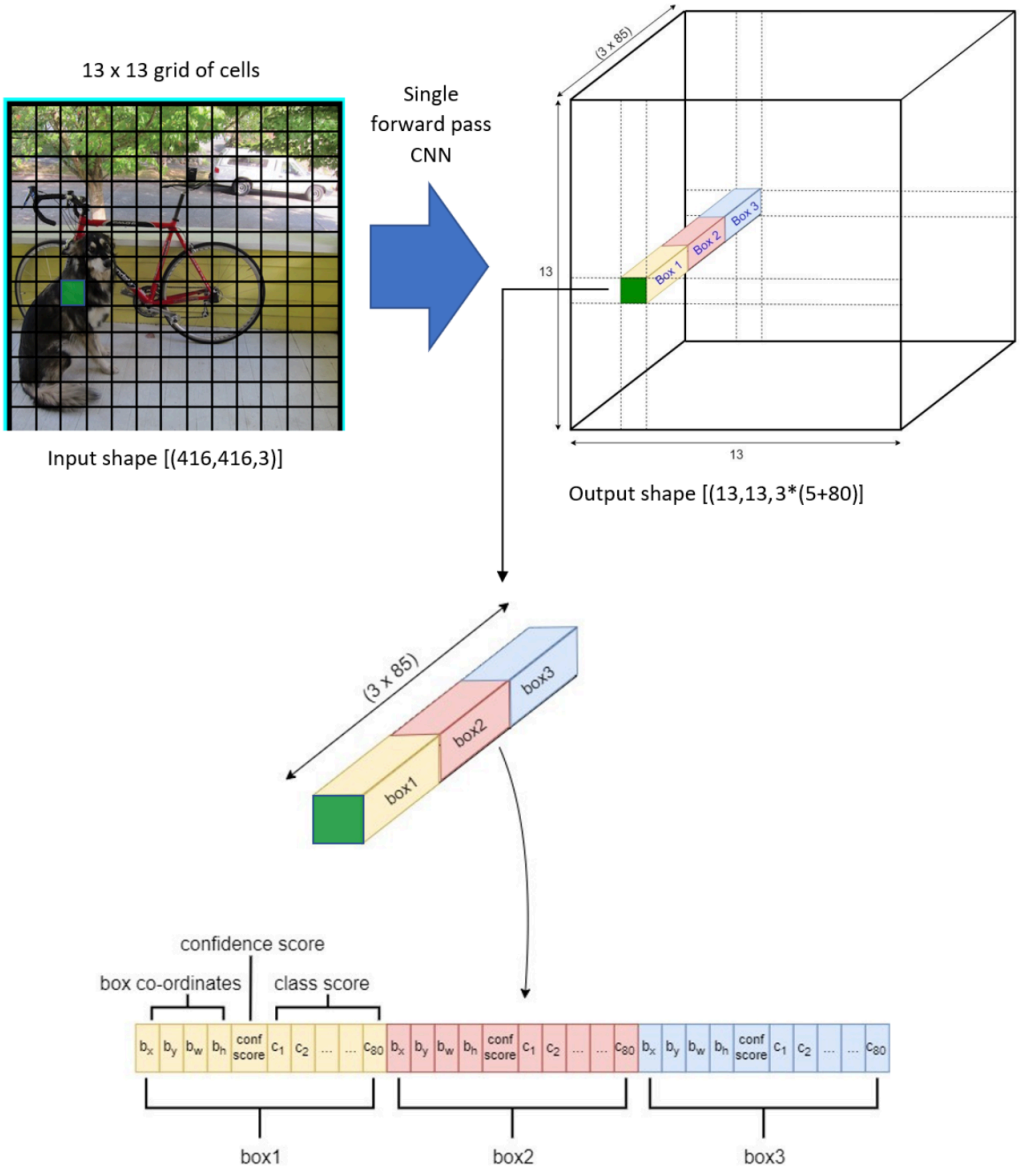
۱.۱ الف

مطابق با متن مقاله شبکه ابتدا عکس را از نظر طول و عرض به یک سایز در می‌آورد، سپس آن را بصورت $S \times S$ تقسیم به S^2 سلول مساوی می‌کند، حال برای هر یک از این سلول ها تعدادی $BoundingBox$ تعریف میشود که تعداد آن ها را با B نمایش می‌دهیم ، هر یک از این باکس ها ۵ مولفه دارند که شامل فاصله طولی و عرضی آنها از مبدا سلول مورد نظر که اصولا سمت چپ بالا انتخاب میشود ، در ورژن یک مقاله، نسبت طول و عرض باکس به طول و عرض عکس و یک اطمینان است که مقدار آن برابر حاصل ضرب احتمال وجود اشیا در این باکس در اشتراک آن با اشیای موجود در سوال است ، که جمعا ۵ مولفه میشود، از طرف دیگر خود سلول نیز یک احتمال اطمینان دارد که بصورت احتمال شرطی $C = P(class_i|Object)$ تعریف میشود ، که برای هر سلول فقط یک مقدار دارد و آن هم بیشینه این احتمال بین کلاس های مختلف است و لذا در نتیجه هر عکس در لایه نهایی تبدیل به یک تنسور میشود که $S \times S \times (B \times 5 + C)$ بعد دارد که در اینجا چون صورت سوال گفته ۳۰۰ کلاس داریم پس $|C| = 300$ است.

۲.۱ ب

ورژن یک این شبکه بر روی داده های با اورلپ عملکرد ضعیفی دارد، چرا ؟چونکه گفته شد برای هر سلول فقط یکی از احتمال های تعریف شده برداشته میشود که آن هم ماکسیمم انها است، این یعنی چه؟ یعنی اگر چند شی در ان سلول باشند، فقط یک شی برداشته میشود و شبکه نمیتواند برای یک سلول بیشتر از یک شی را دیتکت کند، این مشکل از روی ساختار شبکه تیز واضح است، چرا که فقط یک لایه دنس فولی کانکتد در آن قرار داده شده که وظیفه آن ست کردن حواشی باندینگ باکس با استفاده از رگرشن است که این باندینگ باکس فقط میتواند روی یک دیتا عملکرد خوبی داشته باشد و در صورت زیاد شدن دیتاها لایه دنس سردرگم میشود که باید رگرسیون باندینگ باکس را برای کدام یک از ابجکت های مختلف انجام بدهد و این یک ضعف است، راه حل ابتدایی که به ذخن میرسد این است که چندین لایه دنس برای چندین رگرسور مختلف متناظر با چندین شی در ساختار شبکه قرار داده شود و اینگونه برای یک سلول فقط یک احتمال یک ابجکت در نظر گرفته نشود.

در ساختار ورژن دوم برای جلوگیری از این خطا در وهله اول لایه های دنس برداشته شده اند و احتمال شرطی گفته شده بجای محاسبه شدن برای هر سلول برای هر باندینگ باکس محاسبه میشود ، نتیجتا میتوان برای سلول های مختلف شی های مختلف نیز پیش بینی کرد که معادل بصری آن را در عکس مشاهده میکنید.



این کار سبب شد که بتوان برای باندینگ های مختلف کلاس های مختلف پیش بینی کرد که چون باندینگ باکس ها متناظر سلول بودند پس برای یک سلول میتوان کلاس های مختلف همزمان پیش بینی کرد، ولی !! اگر شی های مختلف درون سلول همچنان از یک نوع شی باشند، الگوریتم باز هم به مشکل میخورد که راه حل آن در ورژن سوم گفته شده و پیاده سازی شده است.

۳.۱ ج

برای جلوگیری از این شکل تا جایی که من فهمیدم دو الگوریتم اتخاذ شده است که به شرح زیر هستند .

- در ابتدا چون شبکه ساختار بسیار عمیق شده است و هرچه جلوتر میرویم رزلوشن کاهش پیدا میکند و باعث میشود اشیای کوکتر گم بشوند، پس شبکه در اول ها، اواسط و اخر شبکه ۳ بار اقدام به پیدا کردن باندینگ باکس ها می‌کند که نتیجه آن این است که میتواند اشیای مختلف درون عکس از سایز کوچک تا بزرگ را باندد کند.
- در وهله دوم بر خلاف ساختارهای قبلی بجای استفاده از تابع فعالیت سافت مکس که کلاس ها را مستقل از هم در نظر میگیرد و نتیجتا اگر کلاسی زیر مجموعه دیگری باشد فقط بزرگتر را انختخاب میکند، نویسنده ها از لاجستیک رگرشن و تعریف یک حد استانه استفاده کرده‌اند که منجر به تشخیص طبف وسیع تری از کلاس ها برای هر سلول شده و وسعت تشخیص الگوریتم را افزایش داده است.