

تولید یک معیار:

برای مرور تا اینجا در ابتدا معیار های مهم برای ارزیابی شبکه های تشخیص عمق و تشخیص اشیا به صورت جداگانه را معرفی و مورد بررسی قرار دادیم و نکات مهم درباره ی آن ها را گفتیم. حال می خواهیم معیار مورد نیاز برای تخمین معیار شبکه را مورد بررسی قرار دهیم. همانطور که در قسمت های قبلی نیز به طور ضمنی بحث کردیم، هنگامی که دو شبکه را به صورت joint استفاده می کنیم، اطلاعات شبکه ها در اکثر مواقع می تواند به یکدیگر کمک کند. اما هنگامی که دو شبکه به صورت سری به هم متصل می شوند، در واقع از لحاظ مفهوم، شبکه اول نقش information دادن به شبکه دوم را دارد و این شبکه دوم است که با توجه به آن که در این حالت نسبت به حالتی که تنها بود، ورودی با اطلاعات بیشتری را دریافت می کند، احتمالاً قادر خواهد بود که خروجی بهتری نیز تولید نماید.

برای مثال اگر شبکه تشخیص عمق اول باشد، عمق هایی که به تصویر می دهد، باعث تشخیص بهتر شبکه تشخیص اشیا خواهد شد. برای درک این موضوع فرض کنیم یک کیف نارنجی با پس زمینه غروب خورشید داشته باشیم. در این حالت برای شبکه تشخیص شی تشخیص درست سخت می شود. در حالی که اگر ابتدا تصویر وارد شبکه تشخیص عمق شود و سپس وارد تشخیص شی شود، این اطلاعات را خواهد داشت که مثلاً خورشید در فاصله دورتری از کیف قرار دارد و در نتیجه می تواند بهتر تشخیص دهد. همانطور که در قسمت های قبلی نیز بحث شد، در حالت سریال برعکس نیز این موارد صادق بوده و مثلاً پیکسل های یک bounding box احتمالاً عمق مشابهی نیز دارند و همین موضوع می تواند باعث بهبود شبکه ی تشخیص عمق شود.

پس به صورت کلی می توان این را گفت که در واقع این معیار شبکه دوم است که معیار اصلی است و معیار شبکه اول که بیشتر نقش تهیه اطلاعات را دارد، اصلاح کننده ی معیار شبکه ی دوم خواهد بود. بنابراین می توان از این مفاهیم برای تولید معیار برای تخمین معیار شبکه joint استفاده کرد. برای مثال در اینجا موردی را که در آن ابتدا شبکه تشخیص شی و سپس شبکه تشخیص عمق قرار دارد، مورد بررسی قرار می دهیم. موارد دیگر نیز کاملاً مشابه همین موضوع می باشد.

در اینجا شبکه obj detection در واقع bounding Box هایی را تشخیص خواهد داد و انگار تصویر ماسک شده را در اختیار شبکه تشخیص عمق قرار می دهد و این یعنی شبکه تشخیص عمق می تواند به صورت ریزبینانه تر تصویر را بررسی کند. (در واقع اگر بخواهیم در مورد منبع این ایده صحبت کنیم، باید گفت از این ایده در بینایی ماشین در قسمت تشخیص تحرک استفاده می شود. در آنجا ابتدا یک تخمین به دست می آید و سپس با استفاده از شبکه هایی به صورت ریزبینانه تر شبکه مورد بررسی قرار می گیرد. مشابه این ایده در مقاله تمرین ۲ نیز آمده بود که ابتدا تخمین و سپس با استفاده از شبکه های cascade شبکه اصلی را به دست می آورد.) در واقع در اینجا شبکه ۱ به شبکه ۲ این امکان را می دهد که دقیق تر بررسی کند. در نتیجه برای مثال اگر معیار کلی را بخواهیم به دست آوریم، اگر فرض کنیم که شبکه ی تشخیص اشیا باعث می شود که یک درصدی از تصویر mask شود و ابعاد تصویر به صورت یک

ضریب کاهش یابد، بنابراین معیار شبکه نیز باید به صورت توانی (لگاریتمی) روی معیار شبکه ۲ اثر بگذارد. برای همین معیار ما برای تخمین تخمین معیار به صورت زیر خواهد بود:

$$(RMSE - \log)_{joint} = (RMSE - \log)_{alone} * \lambda * \log(0.5 + mAP^{1-\alpha} + 0.5^{\alpha})$$

آن چه در این جا اهمیت دارد توجه به آن است که اگر bounding box هایی که یک توزیع از دیتا به ما می دهد، overlap زیادی داشته باشد، احتمالاً منجر به این می شود که شبکه ی دوم بیشتر به خطا بیفتد. برای همین در رابطه یک α تعبیه شده تا این اثر را پیش بینی کند. از آن جا که معیاری که می خواهیم تولید کنیم، باید یک معیار general باشد و چون دیتای لیبل خورده برای هر دو (عمق و شی) را همزمان نداریم و نمی توان با آزمایش ضریب آلفا مناسب برای حالت کلی را تایید کرد و با توجه به آن که هدف سوال تخمین معیار شبکه JOINT از روی شبکه های تشخیص شی و تشخیص عمق است و ما آن شبکه ها را داریم، می توان برای دیتاست NYU که در اختیار داریم، لاندرا میانگین مساحت تمام box های دیتا (به ازای تمامی تصاویر) که با استفاده از شبکه obj detection تنه ایمان به دست آوریم، در نظر بگیریم و با استفاده از اعداد موجود آن را به گونه ای به دست می آوریم که برای دیتاست NYU مناسب باشد. حال هنگامی که دیتاست ما عوض شد، برای به دست آوردن alpha در این معیار آن را به شبکه obj از قبل آموزش داده شده می دهیم و alpha ی به دست آمده از شبکه را (میانگین مساحت تمامی bounding box برای دیتاست جدی) به دست می آوریم و به صورت نرمالایز شده (مپ شده به ۱) در رابطه قرار می دهیم.