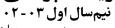
پردازش هوشمند تصاویر زیست پزشکی



مدرس: محمدحسین رهبان



دانشگاه صنعتی شریف دانشکدهی مهندسی کامپیوتر

کوییز پنجم (۲۰ نمره)

1. به سوالات زیر در مورد روشهای object detection پاسخ دهید. (۶ نمره)

• دليل اصلي كند بودن روش R-CNN چيست؟

پاسخ:

در روش R-CNN ابتدا با استفاده از یک روش تعدادی Region به دست می آید. سپس هر کدام از این Region ها به شبکه عصبی داده می شوند تا برای هر کدام فیچرمپ استخراج شود. به همین دلیل به تعداد Region هایی که اول استخراج کرده ایم باید forward pass داشته باشیم که باعث کند شدن این روش می شود.

• دلیل اصلی سریعتر بودن روش Fast R-CNN نسبت به روش قبلی چیست؟ پاسخ:

در این روش ابتدا تصویر اصلی را به شبکه می دهیم و سپس با توجه به region هایی که با روش مشابه روش قبل بدست آمده فیچرمپ های مربوط به region های مختلف بدست می آید و سپس پردازش های بعدی انجام می شود. بنابراین چون هر region به طور جداگانه به شبکه داده نمی شود این روش سریعتر عمل می کند.

• در روش Faster R-CNN چه تغییری در روش قبل داده شده که آن را سریعتر میکند؟ پاسخ:

در این روش برای بدست آوردن region ها هم از شبکه عصبی استفاده میکنیم که به آن region proposal در این روش برای بدست آوردن network گفته می شود. استفاده از این شبکه باعث افزایش سرعت نسبت به روشهای قبل می شود.

- ۲. به سوالات زیر در مورد مبحث interpretability پاسخ دهید. (۹ نمره)
- در مورد اهمیت و ضرورت استفاده از روشهایی برای تفسیر و تحلیل کردن مدل دو مورد را بیان کنید. یاسخ:

• در روشهای بحث شده در کلاس در مبحث attribution methods تابع saliency را به صورت زیر تعریف کردیم:

$$Saliency(x) = \frac{\partial f(x)}{\partial x}$$

چرا هیتمپهای تولید شده در این روش دارای نویز است؟

ياسخ:

۱) ممکن است پیکسلهایی که در تصمیمگیری مدل موثر بوده است به صورت رندومی بخش شده باشد و این نویز در تصمیمگیری مدل اهمیت دارد. ۲) مشکل discontinuous بودن گرادیان Υ) ممکن است saturation point خود رسیده باشد.

• در روش مربوط به sensitivity analysis یکی از روشها محاسبه ی تابع زیر است:

Relevance score =
$$R(x_i) = (\frac{\partial f(x)}{\partial x})^2$$

فرض کنید یک تسک binary classification با ترشولد صفر و ورودیهای دو بعدی داریم. برای تابع داده شده و دو نقطهی x_1, x_2 مقدار خروجی (کلاس پیشبینی شده) و همچنین مقدار relevance score را محاسبه کنید. با توجه به مقادیر به دست آمده، این score چه چیز را نشان می دهد؟ ۱) دلیل پیشبینی لیبل y برای y کنید. با تثیر هر فیچر در اطمینان پیشبینی انجام شده ؟

$$f(x) = x^2 + y^2 - 9$$
, $x_1 = (-1, 1)$, $x_2 = (-4, -4)$

پاسخ:

$$\nabla f(x,y) = [\partial f/\partial x \ \partial f/\partial y] = [2x \ 2y]$$

$$x_1: f(x_1) = -7 < 0, \ classN, \ gradient: [-2,2], \ sensitivity: [4,4]$$

$$x_1: f(x_1) = 7 > 0, \ classP, \ gradient: [-8,-8], \ sensitivity: [64,64]$$

دقت کنید که sensitivity تنها نشان می دهد که با افزایش و یا کاهش x و y ممکن است به نقطهای برسیم که با اطمینان بیشتری در کلاس N پیش بینی میشود.

 $m{7.}$ تصویری با ابعاد ورودی 224×224 پیکسل را در نظر بگیرید. اگر شما از (Vision Transformer (ViT) با patch های غیر همپوشان با ابعاد 16×16 استفاده کنید، چه مرتبه محاسباتی برای پردازش این تصویر لازم است؟ (۵) نمره) **یاسخ:**

طبق ابعاد داده شده، تعداد patch برابر $14 \times 14 = (224/16) \times (224/16)$ است. حالاً باید مرتبه محاسباتی مورد نیاز برای پردازش این تصویر را محاسبه کنیم. زمانی که از ViT استفاده می کنیم، محاسبات بر روی patch ها انجام می شود. بنابراین، تعداد محاسبات مورد نیاز به مرتبه $O(n^2)$ است، که n تعداد patch است. در اینجا، $O(n^2)$ است، $n=14 \times 14=196$ است. بنابراین، مرتبه کلی محاسباتی برابر است با $n=14 \times 14=196$.