پردازش هوشمند تصاویر زیست پزشکی

نیم‌سال اول ۰۳-۰۲

مدرس: محمدحسین رهبان

دانشگاه صنعتی شریف

دانشکده‌ی مهندسی کامپیوتر

کوییز پنجم (۲۰ نمره) | مدت زمان: ۲۵ دقیقه

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

- دلیل اصلی کند بودن روش R-CNN چیست؟

پاسخ:

در روش R-CNN ابتدا با استفاده از یک روش تعدادی Region به دست می‌آید. سپس هر کدام از این Region ها به شبکه عصبی داده می‌شوند 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 می‌شود.

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

پاسخ:

در این روش ابتدا تصویر اصلی را به شبکه می‌دهیم و سپس با توجه به region هایی که با روش مشابه روش قبل بدست آمده فیچرمپ های مربوط به region های مختلف بدست می‌آید و سپس پردازش‌های بعدی انجام می‌شود. بنابراین چون هر region به طور جداگانه به شبکه داده نمی‌شود این روش سریعتر عمل می‌کند.

- 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 network گفته می‌شود. استفاده از این شبکه باعث افزایش سرعت نسبت به روش‌های قبل می‌شود.

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.

پاسخ:

۱) بررسی اینکه مدل به درستی کار می‌کند. ۲) ارتقا دادن، بهتر کردن و برطرف کردن مشکلات مدل

- 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؟

پاسخ:

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 ۳) ممکن است f(x) به saturation point خود رسیده باشد.

- در روش مربوط به sensitivity analysis یکی از روش‌ها محاسبه‌ی 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 می‌دهد؟ ۱) دلیل پیش‌بینی لیبل y برای x و یا ۲) تاثیر 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 می‌دهد که با افزایش و یا کاهش x و y ممکن است به نقطه‌ای برسیم که با اطمینان بیشتری در کلاس N پیش‌بینی میشود.

۳. تصویری با ابعاد ورودی 224×224 پیکسل را در نظر بگیرید. اگر شما از Vision Transformer (ViT) با patchهای غیر همپوشان با ابعاد 16×16 استفاده کنید، چه مرتبه محاسباتی برای پردازش این تصویر لازم است؟ (۵ نمره)

پاسخ:

طبق ابعاد داده شده، تعداد patchها برابر $(224/16) \times (224/16) = 14 \times 14$ است. حالا باید مرتبه محاسباتی مورد نیاز برای پردازش این تصویر را محاسبه کنیم. زمانی که از ViT استفاده می‌کنیم، محاسبات بر روی patch ها انجام می‌شود. بنابراین، تعداد محاسبات مورد نیاز به مرتبه $O(n^2)$ است، که n تعداد patchها است. در اینجا، $n = 14 \times 14 = 196$ است. بنابراین، مرتبه کلی محاسباتی برابر است با $196^2 = 38416$.