



تحلیل هوشمند تصاویر زیست پزشکی

نیم‌سال اول ۰۴-۰۳

مدرس: محمدحسین رهبان

کوئیز پنجم

زمان: ۲۰ دقیقه

۱. مسأله تشخیص تومور در تصاویر X-ray با اندازه 100×100 را در نظر بگیرید. هدف این است که با دریافت یک تصویر ورودی، پیکسل‌های مربوط به مناطق تومور در تصویر خروجی به صورت سفید مشخص شوند. داده‌های آموزشی به صورت زوج‌های تصویر ورودی و ماسکی که محل تومور را مشخص می‌کند، به شکل $\{(x^n, y^n)\}_{n=1}^N$ برای آموزش شبکه در اختیار است.

- شکل شبکه پیشنهادی را مشخص کنید. (همراه با ذکر نوع تمام لایه‌ها)
- نوع لایه خروجی، ابعاد خروجی شبکه، و نحوه محاسبه پیکسل‌های خروجی را به صورت دقیق توضیح دهید.
- نحوه محاسبه تابع هزینه (Loss) را به صورت دقیق و همراه با فرمول ارائه کنید.

پاسخ:

۱. معماری شبکه (Network Architecture)

برای حل مسأله تشخیص تومور، می‌توان از معماری encoder-decoder مانند U-Net استفاده کرد. اجزای اصلی این معماری شامل موارد زیر هستند:

- **Encoder:** وظیفه استخراج ویژگی‌های سلسله‌مراتبی را دارد و از بلوک‌هایی شامل موارد زیر تشکیل می‌شود:
 - لایه‌های Conv2D با تعداد فیلترهای افزایشی.
 - تابع فعال‌سازی (مثلاً ReLU).
 - لایه کاهش ابعاد (مانند MaxPooling) برای کاهش ابعاد فضایی.
- **Bottleneck:** این بخش عمیق‌ترین قسمت شبکه است که ویژگی‌های سطح بالا را استخراج می‌کند و از چندین لایه Conv2D با بیشترین تعداد فیلتر تشکیل شده است.
- **Decoder:** وظیفه بازسازی نقشه‌های ویژگی به ابعاد اصلی تصویر را دارد و شامل مراحل زیر است:
 - لایه UpSampling2D یا Conv2DTranspose برای افزایش ابعاد فضایی.
 - لایه Conv2D برای کاهش تدریجی تعداد فیلترها.
 - اتصال‌های skip connection با لایه‌های متناظر در بخش encoder.
- **خروجی نهایی:** یک لایه Conv2D با یک فیلتر (channel 1) و تابع فعال‌سازی Sigmoid برای تولید احتمالات پیکسل به پیکسل.

۲. خروجی شبکه و نحوه محاسبه آن

- نوع لایه خروجی: آخرین لایه شبکه Conv2D است که از تابع فعال‌سازی Sigmoid استفاده می‌کند.
- ابعاد خروجی: ابعاد خروجی شبکه $H \times W \times 1$ است. در اینجا H و W به ترتیب ارتفاع و عرض تصویر ورودی هستند، و کانال خروجی برابر 1 است که نشان‌دهنده احتمال وجود یا عدم وجود تومور در هر پیکسل است.
- نحوه محاسبه پیکسل‌های خروجی: برای هر پیکسل (i, j) ، مقدار احتمال از طریق اعمال Sigmoid روی مقدار لایه نهایی به دست می‌آید:

$$p_{ij} = \sigma(z_{ij}),$$

که در آن z_{ij} مقدار ویژگی قبل از تابع Sigmoid است.

۳. محاسبه تابع هزینه (Loss Function)

تابع هزینه پیشنهادی Binary Cross-Entropy Loss است که به صورت زیر تعریف می‌شود:

$$\mathcal{L} = -\frac{1}{N \times H \times W} \sum_{n=1}^N \sum_{i=1}^H \sum_{j=1}^W (y_{ij}^n \log(p_{ij}^n) + (1 - y_{ij}^n) \log(1 - p_{ij}^n)),$$

که در آن:

- N : تعداد نمونه‌های آموزشی.
- H, W : ابعاد تصویر ورودی.
- y_{ij}^n : مقدار واقعی ماسک در مختصات (i, j) برای نمونه n .
- p_{ij}^n : احتمال پیش‌بینی شده توسط شبکه در همان مختصات.

نکته تکمیلی

ابعاد خروجی باید شامل یک کانال (1) باشد تا نشان‌دهنده احتمال وجود تومور در هر پیکسل باشد. برای دقت بیشتر روی مناطق کوچک مانند تومور، می‌توان از ترکیب Binary Cross-Entropy Loss با Dice Loss استفاده کرد.

۲. با توجه به مباحث مربوط به SAM به سوالات زیر پاسخ دهید:

- چرا در مدل SAM از سه ماسک در خروجی استفاده میشود؟
- تابع هزینه در این مدل زوی کدام ماسک محاسبه میشود؟
- نحوه آموزش تعاملی این شبکه را شرح دهید.

پاسخ:

- مدل Segment Anything (SAM) سه خروجی ماسک ارائه می‌دهد تا ابهامی که ممکن است از یک input prompt ایجاد شود را مدیریت کند. این سه ماسک معمولاً لایه‌های part، whole و subpart را توصیف می‌کنند که برای نمایش ماسک‌های nested کافی هستند. با پیش‌بینی هم‌زمان چندین ماسک، مدل می‌تواند تفسیرهای مختلف یک prompt را بهتر شناسایی کند و عملکرد segmentation را بهبود بخشد.
- در طول training، مدل loss را بین ground truth و هرکدام از ماسک‌های پیش‌بینی شده محاسبه می‌کند، اما فقط loss مربوط به دقیق‌ترین ماسک را برای backpropagation استفاده می‌کند. این روش باعث می‌شود مدل روی یادگیری دقیق‌ترین ماسک تمرکز کند و همچنان قابلیت تولید خروجی‌های متنوع را برای مدیریت سناریوهای مختلف ambiguity داشته باشد.
- پس از پیش‌بینی اولیه، نقاط بعدی به طور یکنواخت از error region بین پیش‌بینی ماسک و ماسک واقعی انتخاب می‌شوند. هر نقطه به عنوان foreground (false negative) یا background (false positive) طبقه‌بندی می‌شود. ماسک پیش‌بینی از مرحله قبلی به عنوان یک prompt اضافی به مدل داده می‌شود و برای ارائه اطلاعات بیشتر در مرحله بعد، unthresholded mask logits به جای ماسک باینری شده استفاده می‌شود. وقتی چندین ماسک بازگشت داده می‌شود، ماسکی که پیش‌بینی IoU بالاتری دارد برای مرحله بعدی انتخاب می‌شود تا نقاط بعدی نمونه‌برداری شوند.