

تحلیل هوشمند تصاویر پزشکی

تمرین دوم، قسمت تئوری

جواد راضی (۴۰۱۲۰۴۳۵۴)

سوال یک

قسمت الف)

فرض کنیم ماتریس تصویر بازسازی شده اندیس‌هایی مانند ماتریس زیر داشته باشد:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

با در نظر گرفتن مقادیر پراجکشن از زوایای مختلف، می‌توان معادلاتی تشکیل داده و این ۴ مقدار را پیدا کرد.

برای زاویه ۰ درجه داریم:

در زاویه ۰، پرتو پراجکشن از دو اندیس a, b می‌گذرد. مقداری که دریافت کرده‌ایم ۱۰ است، که می‌توان گفت حاصل

جمع $Intensity$ در این نقاط بوده‌است. پس داریم: $a + b = 10$

با همین منطق، در $r = 1$ ، داریم: $c + d = 7$

در زاویه ۴۵ درجه، اگر پرتو پراجکشن را از سمت چپ در نظر بگیریم، در $r = 0$ ، این پرتو فقط از a عبور می‌کند. با

توجه به اینکه مقدار دریافتی ۷ است، پس داریم:

$$a = 7 \rightarrow$$

$$b = 10 - a = 3$$

در زاویه ۹۰ درجه، پراجکشن مشابه زاویه ۰ است، اما به صورت عمودی. در فاصله صفر، داریم:

$$a + c = 12 \rightarrow c = 12 - a = 5$$

برای فاصله ۱ $r = 1$ نیز رابطه زیر را داریم:

$$b + d = 5 \rightarrow d = 5 - b = 2$$

اکنون تمام مقادیر را یافتیم و نیازی به بررسی بقیه پراجکشن‌ها نیست. با این مقادیر، ماتریس به صورت زیر بازسازی می‌شود:

$$\begin{bmatrix} 7 & 3 \\ 5 & 2 \end{bmatrix}$$

قسمت ب)

در CT-Scan، واحد Hounsfield، نمایانگر Radio Intensity بافت در هر محل است. مقدار HF هوا ۱۰۰۰-، و برای آب ۰ است. با توجه به اینکه دقیقاً در صورت سوال گفته نشده اطلاعات تصویر بازیافت‌شده به چه صورت است و با چه اسکالی می‌توان آن‌ها را به HF تبدیل کرد، می‌توان از روی بزرگی اعداد حدس‌هایی زد:

ستون ۷،۵ احتمالاً مربوط به بافت dense تر، یعنی استخوان است.

ستون ۳،۲، مربوط به جایی است که density پایین‌تری دارد و احتمالاً مربوط به هوا می‌باشد.

قسمت ج)

- سرعت: عملیات Back Projection، از نظر محاسباتی سنگین است؛ برای بازسازی هر پیکسل، احتمالاً باید از تمام زوایایی که پراجکشن دریافت‌شده است، از روی مقادیر انتگرال گرفته و پیکسل رو محاسبه کنیم. اگر این کار برای پیکسل به پیکسل عکس انجام شود، پیچیدگی محاسباتی قابل توجهی خواهد داشت.
- حساسیت به نویز: این روش به نویز حساس است و خطاهای کوچک در اندازه‌گیری، می‌توانند منجر به ایجاد آرتیفکت‌هایی در تصویر شوند.
- نیاز به حافظه: در این روش، نیاز است که دیتای پراجکشن را برای تمام زوایا در حافظه ذخیره کنیم. این کار، حافظه زیادی می‌طلبد. خصوصاً در تصاویر سه‌بعدی که از آبجکت‌های با اندازه نسبتاً بزرگ گرفته می‌شود، مقدار حافظه مورد نیاز بسیار بالاست.

سوال دوم)

قسمت الف)

در MRI، عموماً کانترست از تفاوت بافت‌ها، در زمان‌های T1 Relaxation و T2 Relaxation می‌آید. نخستین زمان، بیانگر این است که چقدر طول می‌کشد تا برآیند میدان مغناطیسی اتم‌های بافت، پس از پالس رادیویی به حالت عادی بازگردد. (در واقع، زمان ریکاوری اتم‌های بافت). دومین زمان، بیانگر زمان میرایی پالس رادیویی دریافتی است، که این مقدار نیز برای بافت‌ها متفاوت است. با تنظیم پارامترهای بهینه توسط اپراتور، ایجاد کانترست در تصویر MRI، به صورتی که برای شناسایی و دیدن جزئیات بافت هدف بهینه باشد، مقدور است.

قسمت ب)

مطابق اسلایدها، سیگنال بدست آمده، دارای یک مولفه سینوسی می‌باشد که سرعت زاویه‌ای آن (سرعت چرخش)، قابل تنظیم است:

$$S(t) = e^{-\frac{t}{T_2}} \cos(\omega t)$$

$$\text{If we set } \omega = 0, \text{ we have: } S(t) = e^{-\frac{t}{T_2}}$$

بدین ترتیب، با صفر کردن سرعت زاویه‌ای، مولفه سینوسی حذف شده و دیتایی که داریم شدت و میرایی سیگنال را نشان خواهد داد.

قسمت ج)

یافتن فرکانس:

مطابق اسلایدهای درس، این روابط را برای فرکانس، نرخ gyromagnetic، و میدان مغناطیسی داریم:

$$f = \gamma B$$

$$B = 2.6 + 0.3 \cdot z : \text{ Given } z = 2 \rightarrow B = 2.6 + 0.6 = 3.2 \text{ T}$$

$$\gamma: \text{ Gyromagnetic Ratio of Hydrogen} = 42.5 \frac{\text{MHz}}{\text{T}}$$

$$\rightarrow f = 42.5 \frac{\text{MHz}}{\text{T}} * 3.2 \text{ T} = 136 \text{ MHz}$$

یافتن زاویه چرخش (Flip Angle):

مطابق اسلایدهای درس، این رابطه را برای یافتن زاویه چرخش داریم:

$$S(t) = \sin(\theta) \gamma B \cos(\omega t) \rightarrow \max S(t): \omega = 0 \rightarrow$$

$$\max S(t) = 68 * 10^7 \text{ MHz} \rightarrow$$

$$68 * 10^7 = 10^7 * \sin(\theta) * 42.5 * 3.2 \rightarrow \sin(\theta) = \frac{68}{42.5 * 3.2} = \frac{1}{2} \rightarrow$$

$$\theta = \text{Flip Angle} = \arcsin\left(\frac{1}{2}\right) = 30^\circ = \frac{\pi}{6}$$

سوال سوم

سه ماده به صورت متوالی قرار گرفته‌اند. برای هر کدام، خروجی را حساب کرده و آن را به عنوان ورودی ماده بعد در نظر می‌گیریم. از سمت چپ شروع می‌کنیم:

$$n_1 = n_0 e^{-\int_0^{l_1} 2x^2 dx} = n_0 e^{-\frac{2}{3}l_1^3} \rightarrow$$

$$n_2 = n_1 e^{-\int_0^{l_2} (x+3) dx} = n_0 e^{-\frac{2}{3}l_1^3} * e^{-\frac{1}{2}l_2^2 - 3l_2} \rightarrow$$

$$n_3 = n_2 e^{-\int_0^{l_3} \frac{1}{3} dx} = \left[n_0 e^{-\frac{2}{3}l_1^3} * e^{-\frac{1}{2}l_2^2 - 3l_2} \right] * e^{-\frac{1}{3}l_3} \rightarrow$$

$$\text{Output} = n_3 = n_0 e^{-\frac{2}{3}l_1^3 - \frac{1}{2}l_2^2 - 3l_2 - \frac{1}{3}l_3}$$

سوال چهارم

قسمت ۱)

مزایای سیستم‌های دیجیتالی PET-CT:

- ریکاوری سیگنال در آن‌ها بهبود یافته‌تر است. این امر تشخیص‌پذیری ضایعه‌های کوچک را نیز امکان‌پذیر می‌کند.

- استفاده این سیستم‌ها از سیلیکون-فوتومالتیپلایر (SiPM)، منجر به افزایش حساسیت **Detector** می‌شود که پس‌آمد این امر، بالا رفتن رزولوشن فضایی تصویر است.
- این سیستم‌ها در مجموع زمان کم‌تر، و دخالت انسانی کم‌تری برای تصویربرداری نیاز دارند، در عین حال که به کیفیت تصویر لطمه نمی‌زنند.

قسمت ۲)

ایده و متد استفاده‌شده:

- ایده این مقاله، تولید تصاویر مصنوعی (**PET Full Dose (Synthetic)**، با استفاده از مجموعه‌ای از تصاویر **Extreme-PET** است که طی ۳۰ ثانیه، با سیستم **PET-CT** اخذ شده‌اند. طبعاً تصویر **PET Full Dose** اگر به صورت واقعی گرفته شود، بیمار را به دوز بالایی از مواد رادیواکتیو **Expose** می‌کند، و در عین حال این تصویربرداری زمان زیادی می‌طلبد و با توجه به نیمه‌عمر بالای مواد رادیواکتیو، پس از تصویربرداری نیز مواد رادیواکتیو در بدن بیمار وجود خواهند داشت که این موضوع برای اطرافیان فرد نیز خطرناک است. با تولید تصاویر مصنوعی از روی تصاویری که زمان کمی می‌برند، امن‌تر هستند و رزولوشن مطلوبی دارند، این ریسک‌ها از بین می‌رود.
- ورودی شبکه عصبی پیشنهادی مقاله، یک تصویر ۶ کاناله است که تصاویر **ExtremePET** و **CT** را که هرکدام در فرمت ۲.۵ بعدی هستند، که با استک‌کردن سه محور مجاور برای هر اسلایس، تصاویر ۳ کاناله ایجاد شده‌اند.
- معماری شبکه عصبی، بر مبنای شبکه **pix2pixHD¹** می‌باشد. این معماری، که نوعی از شبکه‌های **GAN** به حساب می‌آید، شامل دو **generator**، و یک **discriminator** می‌باشد. دو جنریتور، با رزولوشن‌های مختلف تلاش می‌کنند که ویژگی‌های محلی، و گلوبال تصویر را تجمیع کنند. **Discriminator**، که شامل سه واحد جدا است که بر روی تصاویر با اسکیل متفاوت کار می‌کنند، تصاویر تولیدشده را دسته‌بندی می‌کند.

¹ [NVIDIA/pix2pixHD: Synthesizing and manipulating 2048x1024 images with conditional GANs](https://github.com/NVIDIA/pix2pixHD) (github.com)

- سنجه‌های ارزیابی مدل، شامل «سنجه شاخص شباهت ساختاری» (SSIM)، «پیک نرخ سیگنال به نویز» (PSNR)، و خطای MAE بر پایه SUV می‌باشند. این سنجه‌ها، شباهت میان تصویر ساختگی، و تصویر واقعی Full Dose PET را می‌سنجند.
- نتایج مقاله نشان داد که شبکه پیاده‌سازی شده، قادر به تولید تصاویر PET با کیفیت بالا می‌باشد که می‌تواند در یافتن جراحات، و اندازه‌گیری‌های مربوط به آن‌ها کارا باشد. با این حال، مدل ارائه شده در جراحات بزرگ‌تر، عملکرد بهتری نسبت به جراحات کوچک‌تر دارد.

سوال ششم)

قسمت الف)

تصاویر پزشکی، عمدتاً به خاطر عواملی نظیر مدالیتی تصویر، حرکت بیمار در طی تصویر برداری، تداخلات محیطی و ... نویزی هستند. این نویزها، می‌توانند به عنوان حملات تخاصمی بر روی تصویر ورودی به حساب بیایند به عملکرد، و پایداری مدل شبکه عصبی در تخمین ضربه می‌زند. در مقاله به طور خاص به چندین نویز که **Multiplicative** هستند، از جمله نویز **Speckle**، و نویز پواسونی (کوانتومی) اشاره شده است که در تصویربرداری‌های نظیر **X-Ray** و **Ultrasound** وجود دارند. وجود این نویزها، به علت ذات **Multiplicative** آن‌ها، باعث می‌شود فیلترهای خطی که برای **denoising** طراحی شده‌اند بر روی آن‌ها بی‌اثر باشند.

قسمت ب)

چالش‌ها:

- کمیاب بودن تصاویر **annotate** شده پزشکی، که چالشی برای ایجاد لیبل برای مدل، و یادگیری شبکه‌های عصبی عمیق می‌باشد.
- عدم تعادل میان کلاس‌های یک دیتاست، که می‌تواند منجر به ایجاد بایاس در تخمین مدل شبکه عصبی شود. همچنین در صورتی که دیتای آموزش شبکه، توزیع متفاوتی با دیتای واقعی که شبکه با آن تست می‌شود داشته باشد، واریانس مدل بالا خواهد رفت.

- مسائل مربوط به رعایت حریم خصوصی بیماران، و ملاحظات اخلاقی، سبب می‌شود که نتوان تصاویر پزشکی و اطلاعات همراه آن را به صورت عمومی، و متن‌باز، در اختیار همگان قرار داد. مقرراتی که ارگان‌های جهانی، و دولت‌های مختلف در این خصوص وضع کرده‌اند، دسترسی به دیتای کافی را دشوار می‌سازد.

راه‌حل‌ها:

- استفاده از راه‌کارهای **Data Augmentation**، برای تولید تصاویر تصنعی. این راه‌کارها منجر به افزایش حجم دیتاست، تنوع آن، و تولید تصاویر جدید همزمان با حفظ ویژگی‌های تصویر اصلی می‌شود.
- تکنیک‌های تطبیق با دامنه که این امکان را میسر می‌کنند که دیتای لیبل‌دار، از یک دامین مشابه را استفاده کرده و با تکنیک‌هایی نظیر **transfer learning**، برای دیتاستی که لیبل ندارد، لیبل ایجاد کرد. این تکنیک‌ها می‌توانند در حل مشکلات کمبود دیتا، و عدم تعادل کلاس‌ها به کار بیایند.
- استفاده از متدهایی که در عین حفظ محرمانگی داده، و پوشاندن اطلاعات هویتی بیمار، امکان اشتراک آن‌ها را، به گونه‌ای که هویت بیمار ناشناس بماند فراهم می‌کنند. این تکنیک‌ها چند دسته دارند؛ یکی از آن‌ها، روش‌های **Differential Privacy** است که میزانی نوین کنترل‌شده به داده اضافه می‌کند، با این هدف که در عین حال که داده بتواند برای یادگیری شبکه‌های عصبی کارا باشد، نتوان با آن به اطلاعات هویتی بیمار پی برد. تکنیک‌های دیگر نیز شامل حذف اطلاعات هویتی بیماران و جایگذاری آن‌ها با اطلاعات و اسامی ساختگی است.

قسمت ج)

معیارهای عددی، شهود کافی را در خصوص اینکه مدل طراحی شده چگونه کار می‌کند، و چرا کار می‌کند به ما نمی‌دهند. این سنجه‌ها که شامل سنجه‌هایی نظیر **f1, precision, recall, accuracy** می‌باشند، عملکرد مدل را روی دیتاست بررسی می‌کنند، اما ایده‌ای به ما در خصوص منطق عملکرد مدل، ویژگی‌هایی که در هر لایه یادگرفته شده است، و بایاس‌های احتمالی که مدل دارد نمی‌دهند. تکنیک‌های توضیح‌پذیری تصویری، که شامل تصویرسازی‌هایی نظیر هیت‌مپ، یا نقشه اکتیویتی هستند، شهود بهتری در خصوص پروسه تصمیم‌گیری مدل، و سنجش عملکرد آن در تخمین یا طبقه‌بندی می‌دهند. چند مثال از الگوریتم‌های توضیح‌پذیری **Visual**:

- **(CAM) Class Activation Mapping**

- (Grad-CAM) Gradient-weighted CAM
- (LIME) Local Interpretable Model Agnostic Explanation

قسمت د)

تکنیک‌های model-agnostic، نظیر LIME، همان‌طور که از نام‌شان پیداست، دامنه مسئله را در نظر نمی‌گیرند و مستقل از آن هستند. این امر باعث می‌شود این تکنیک‌ها، طیف وسیعی از کارکردها را داشته باشند. از سوی دیگر، این رویکرد که مستقل از مدل است، باعث می‌شود این روش‌ها جزئیات، و پیچیدگی‌های مختص مدل و مسئله را کپیچر نکنند و در کانتکس‌های خاص، دقت کم‌تری داشته باشند.

تکنیک‌های model-specific، نظیر CAM و Grad-CAM، مکانیزم‌های درونی مدل را بررسی و استفاده می‌کنند تا هیت‌مپ‌هایی تفسیرپذیر، که خاص مدل هستند را تولید کنند. این موضوع، نقص تکنیک‌های model-agnostic را رفع می‌کند، اما در عین حال سبب می‌شود پیاده‌سازی این تکنیک‌ها پیچیده بوده، و طیف کاربرد آن‌ها بسیار محدودتر از روش‌های model-agnostic باشد.