



## پردازش هوشمند تصاویر زیست پزشکی

نیم سال اول ۰۳-۰۲

مدرس: محمدحسین رهبان

پاسخ تمرین دوم

مدالیه های تصویربرداری

بخش تئوری

### ۱. سوال اول

الف. چهار متغیر  $x_{11}, x_{12}, x_{21}, x_{22}$  داریم که از ماتریس بدست می آیند.  $x_{ij}$  نمایانگر سطر  $i$  و ستون  $j$  است.

$$\theta = 45 : x_{21} = 7, x_{12} = 2$$

$$\theta = 135 : x_{11} = 5, x_{22} = 3$$

- توجه: از بقیه مقادیر داده شده نیز میتوان بدست آورد، ولی به محاسبات بیشتری نیاز است.
- ب. نقاط با روشنایی بیشتر مربوط به استخوان و با روشنایی کمتر مربوط به هوا هستند، پس ۵ و ۷ استخوان و ۲ و ۳ هوا هستند.
- ج. محاسبات بالایی دارد، به طوری که برای یک تصویر  $n \times n$  باید یک دستگاه  $n$  معادله  $n$  مجهول را حل کرد و اینکار به حافظه زیادی هم نیاز دارد.
- همچنین در عمل تناظر دادن یک پیکسل از تصویر به یک بخش از سلول کار سختی است.
- مورد دیگر هم این است که تصاویر blurry میشوند زیرا نویز هایی که روی تصاویر در زوایای مختلف می افتد با هم متفاوت هستند و حل دستگاه را نادقیق میکند.

### ۲. سوال دوم

- الف. contrast ایجاد شده بین بافت ها به دلیل اختلاف تعداد هیدروژن های آنها میباشد. این اختلاف در تصویر برداری بدین شکل تاثیر میگذارد.
- تغییر زمان  $T_1$ : زمان  $T_1$  مدت زمانی است که یک مولکول برای بازگشت به حالت اولیه خود پس از قرار گرفتن در معرض یک میدان مغناطیسی گذرا طول می کشد. بافت های مختلف دارای زمان های  $T_1$  متفاوتی هستند. به عنوان مثال، آب دارای زمان  $T_1$  کوتاه است، در حالی که چربی دارای زمان  $T_1$  طولانی است. با استفاده از تکنیکی به نام زمان بندی  $T_1$  می توان کنتراست بین بافت های مختلف را بر اساس زمان  $T_1$  آنها ایجاد کرد.
- تغییر زمان  $T_2$ : زمان  $T_2$  مدت زمانی است که یک مولکول برای از دست دادن انرژی خود پس از قرار گرفتن در معرض یک میدان مغناطیسی گذرا طول می کشد. بافت های مختلف دارای زمان های  $T_2$  متفاوتی هستند. به عنوان مثال، چربی دارای زمان  $T_2$  کوتاه است، در حالی که آب دارای زمان  $T_2$  طولانی است. با استفاده از تکنیکی به نام زمان بندی  $T_2$  می توان کنتراست بین بافت های مختلف را بر اساس زمان  $T_2$  آنها ایجاد کرد.
- ب. با مشاهده کردن قله های سیگنال، تاثیر بخش سینوسی حذف میشود و فقط بخش نمایی باقی میماند.
- ج. برای بدست آوردن فرکانس داریم:

$$B = 2.6 + 0.6 = 3.2T$$

$$\omega = \gamma * B = 3.2 * 42.5 = 136MHz$$

و همچنین برای flip angle داریم:

$$s(t) = N * \sin\theta * \gamma * B * \cos(\omega t)$$

زمانی ماکسیمم است که  $\cos(\omega t)$  برابر با یک باشد.

$$68 * 10^7 = 10^7 * \sin(\theta) * 42.5 * 3.2$$

$$\sin(\theta) = \frac{1}{2}$$

$$\theta = 30$$

### ۳. سوال سوم

$$\begin{aligned} N_3 &= N_0 * e^{-\int_0^{L_1} \mu_1(x) dx} * e^{-\int_0^{L_2} \mu_2(x) dx} * e^{-\int_0^{L_3} \mu_3(x) dx} \\ &= N_0 * e^{-\frac{2}{3}L_1^3} * e^{-(\frac{L_2^2}{2} + 3L_2)} * e^{-\arctan(L_3)} \end{aligned}$$

### ۴. سوال چهارم

سیستم های PET/CT دیجیتال از SiPM (silicon-photomultipliers) برای دیتکت کردن تابش ها استفاده میکنند که در مقایسه با دیکتورهای سیستم های قدیمی تر از حساسیت بالاتری برخوردارند و میتوانند تابش ها با ذرات کمتر را نیز detect کنند که این ویژگی موجب میشود در زمان کمتر بتوانیم تابش های بیشتری را detect کنیم و در نتیجه در زمان کمتر اطلاعات بیشتری را بتوانیم دریافت و بازسازی کنیم. این تکنولوژی میتواند زمان تصویر برداری را به حدود یک سوم کاهش دهد و تصاویر حاصل از آن بدون نیاز به پردازش بیشتر، کیفیت کافی برای بررسی های مورد نیاز را داشته باشند.

#### ایده ی کلی:

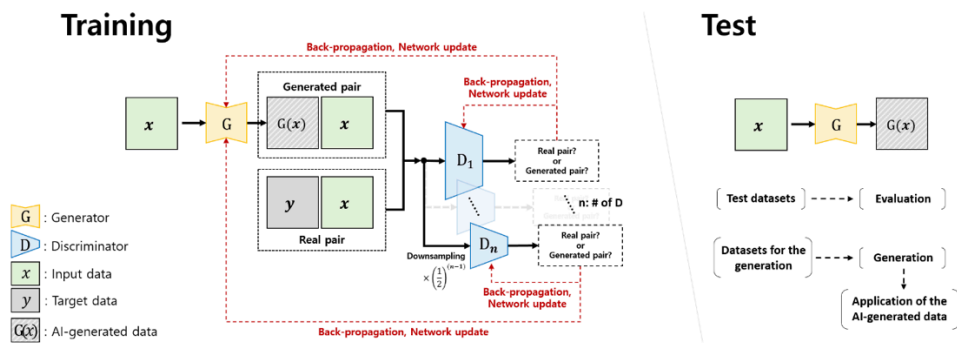
در این مطالعه از بیماران دو تصویربرداری انجام شده. اول PET/CT در زمان عادی که دو نوع تصویر CT scan و FullTime-PET از آن به دست می آید و سپس انجام اسکن PET به صورت low dose و در زمان کمتر که در این حالت تصویر به دست آمده از جزییات و کیفیت کافی برخوردار نیست ( Extreme PET ) و سپس تلاش برای به دست آوردن تصویر با جزییات کافی از روی Extreme PET با استفاده از یک معماری مبتنی بر GAN و همچنین استفاده از تصاویر CT ( بازسازی FullTime PET از روی Extreme PET ).

ورودی: ورودی شبکه در دو حالت فقط تصاویر Extreme PET و Extreme-PET+CT آزمایش شده است. برای بازسازی اسلایس i ام از حجم عکس برداری شده، اسلایس های i-1 و i+1 از حجم هم به ورودی شبکه داده میشود. بنابراین برای حالت اول یک ورودی سه کاناله و برای حالت دوم دو ورودی سه کاناله خواهیم داشت (یک CT و یک PET هرکدام سه اسلایس) طبیعتاً تمام تصاویر نرمالایز و به اندازه ی یکسان scale میشوند. خروجی شبکه هم حالت بهبود یافته ی ExtremePET است (AI-ExtremePET).

نکته ی قابل توجه درمورد تصاویر PET این هست که روشنایی/تاریکی بخش های تصویر بیانگر معیاری به نام SUV هست که میزان فعالیت FDG در بافت عکس برداری شده را مشخص میکند و مقدار آن معیاری برای وجود ضایعات مختلف در بافت ها به ما میدهد. در سنجش بازسازی تصاویر Full time از روی تصاویر Extreme مقدار درست SUV میتواند معیار خوبی باشد.

معماری شبکه :

در این مطالعه از مدل pix2pixHD 2 استفاده شده که معماری آن به صورت زیر است:



Generator در این مدل باید سعی کند از روی ورودی  $x$  (ExtremePET+CT) خروجی  $G(x)$  (پت اسکن با کیفیت) بسازد، و Discriminator باید تلاش کند با دریافت دو جفت تصویر  $(x, G(x))$  و  $(x, y)$  که در واقع FullTime-PET است، تشخیص دهد کدام جفت تصویر متعلق به خروجی Generator است. نکته‌ی قابل توجه در معماری pix 2 pixHD این است که می‌تواند  $n$  Discriminator برای  $n$  اسکیل متفاوت تصویر داشته باشد و در هر اسکیل تشخیص را انجام دهد. همچنین در این معماری دو Generator داریم که در دو رزولوشن مختلف تصویر عمل می‌کنند تا بتوانند در ویژگی‌های محلی و گلوبال تصاویر را تجمیع کنند. تغییر قابل توجهی که در این مقاله روی معماری اصلی pix 2 pixHD انجام شده، افزودن امکان convolution group در اولین لایه‌ی کانولوشن انکودر Generator هاست که به ما این امکان را می‌دهد که برای مواقعی که همزمان تصاویر PET و CT را به عنوان ورودی می‌دهیم، مدل را به گونه‌ای آموزش دهیم که تصاویر PET و CT را به طور جداگانه با استفاده از ۲ گروه با ۳ کانال پردازش کند. در این مقاله سه مدل آموزش داده شده و بهترین آن‌ها با استفاده از معیارهایی که در ادامه توضیح داده می‌شود، انتخاب شده.

مدل M۱: ورودی فقط Extreme PET

مدل M۲: ورودی Extreme PET و CT بدون امکان group convolution

مدل M۳: ورودی Extreme PET و CT با group convolution

## ارزیابی:

در قدم اول مدل با استفاده از معیارهای (SSIM) Structural Similarity Index Measure و (PSNR) Peak Signal to Noise ratio (MAE) SUV-based Mean Absolute Error و (MAE) SUV-based Mean Absolute Error سنجیده شده است. نکته‌ی مهم این است که برای محاسبه این معیارها، تصاویر به گونه‌ای mask شده‌اند که فقط نواحی حائز اهمیت (نواحی بافت/بدن بیمار) در محاسبه دخیل باشند و سایر بخش‌های تصویر mask 0 داشته باشند. در قدم بعدی تعدادی از نمونه‌ها توسط متخصص انسانی مورد بررسی قرار گرفته‌اند. به این صورت که متخصص انسانی برای تصاویر FullTime PET و خروجی بهبود یافته‌ی ExtremePET (AI-ExtremePET) ضایعات ممکن را شناسایی می‌کند و با مقایسه نتایج به دست آمده از تصاویر FullTime و AI-Extreme بررسی می‌شود که چند درصد از ضایعات شناسایی شده در تصاویر FullTime در تصاویر AI-Extreme نیز قابل تشخیص هستند. این بررسی به تفکیک بافت‌ها نیز انجام شده است (یعنی مثلاً چند درصد از ضایعات ریه/استخوان/کبد... از تصویر بازسازی شده قابل تشخیص بوده است). همچنین این موضوع به تفکیک PERCIST-measurable بودن/نبودن ضایعات نیز انجام شده.

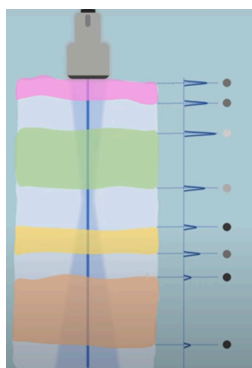
$$\text{PERCIST-measurable: } (SUV_{\text{peak}} > 1.5 \times \text{mean}(SUV_{\text{liver}}) + 2 \times \text{std}(SUV_{\text{liver}}))$$

نکته‌ی مهم این است که برای جلوگیری از بایاس در شناسایی ضایعات در تصاویر FullTime و AI-Extreme متناظر، این تصاویر به طور جدا و رندوم در اختیار متخصصان قرار گرفته باشند. در نهایت، عملکرد مدل M۳ بهترین بوده است.

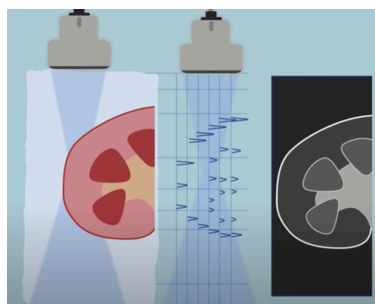
## ۵. سوال پنجم

**A-mode (Amplitude):** ساده‌ترین حالت عکس‌برداری ultrasound است که در آن transducer فقط روی

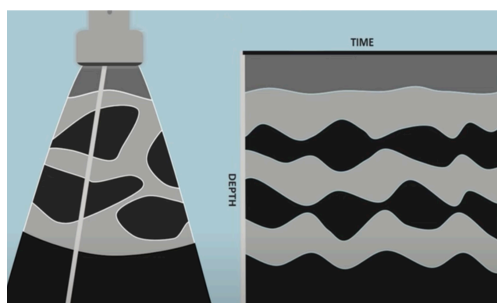
یک محور بازتاب امواج ارسالی در tissue interface ها را دریافت میکند و به این وسیله عمق بافت‌های مختلف در راستای آن محور را اندازه‌گیری می‌کند. کاربرد این مد برای محاسبه‌ی عمق و کشف کیست و تومورهاست.



**B-mode (Brightness)** : متداول‌ترین حالت عکس‌برداری ultrasound است که در آن یک تصویر دو بعدی به دست می‌آید. فرآیند این حالت به این شکل است که عمق و شدت بازتاب امواج از بافت‌ها روی محورهای موازی اندازه‌گیری می‌شود و با در کنار هم قرار دادن آن‌ها به تصویری دو بعدی از سطح عکس‌برداری شده می‌رسیم. کاربرد این مد تصویربرداری دو بعدی از اندام‌هاست و تفسیر آن ساده است.



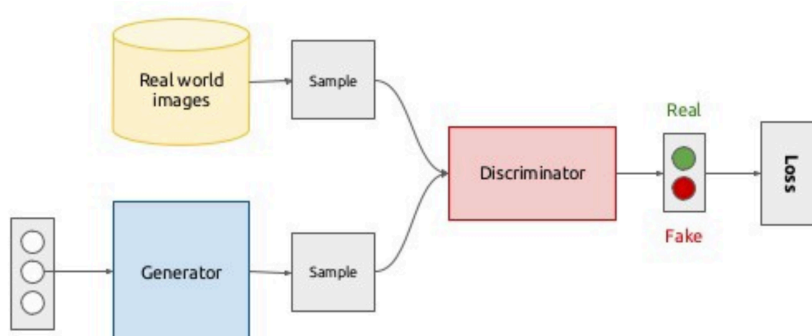
**M-mode (Motion)** : کارکرد این مد مانند یک ویدیوی stop-motion است. به این صورت که مجموعه‌ای از تصویربرداری حالت A در یک محور ثابت (یا B در یک سطح ثابت) در طی زمان‌های پیاپی گرفته می‌شود و با در کنار هم قرار دادن آن‌ها می‌توانیم چگونگی تغییرات و حرکت در آن محور یا سطح را ببینیم (مثل فریم‌های مختلف یک ویدیو که پشت سر هم قرار می‌گیرند). کاربرد این مد برای زیر نظر قرار دادن تغییرات است، مثلاً بررسی تغییرات اندازه‌ی دریچه‌ی قلب در طی زمان.



آ) اگر ورودی شبکه‌های عصبی نویزی باشد و signal-to-noise ratio بالایی داشته باشند، شبکه مستعد - Atten-uation و خطا در پیش‌بینی خواهد بود. تصاویر پزشکی به دلیل چالش‌هایی که در تصویربرداری با آن روبرو هستند (مثلا بازسازی تصاویر از روی حوزه فرکانس و...) مستعد نویز زیادی هستند. اگر به هر دلیلی نویز تصاویر بیشتر از مقدار خاصی باشد یا شبکه تصویری را ببیند که قبلا نویز مشابه نویز آن تصویر را در آموزش ندیده باشد مستعد خطای زیادی خواهد بود که به دلیل حساس بودن حوزه تصاویر پزشکی این موضوع قابل قبول نیست. به دلیل ذات نویز های تصاویر پزشکی پیش پردازش برای از بین بردن آنها نیز کار ساده ای نیست و در خیلی از مواقع موجب از بین رفتن جزئیات مهم تصویر خواهد شد. به همین دلیل مهم است برای جلوگیری از پیش آمدن چنین مشکلاتی مطمئن شویم شبکه درحین آموزش تصاویر نویزی نیز دیده باشد. به این منظور میتوانیم حتی نویزهای متفاوت را به تصاویر اضافه کنیم و شبکه را با آن آموزش دهیم تا شبکه نسبت به نویز مقاوم شود.

ب) برای داشتن مدل های قابل اعتماد طبیعتا برای ما مطلوب است که دیتاسیت های بزرگ و annotated از تصاویر پزشکی داشته باشیم اما جمع آوری چنین دیتاستی به دو دلیل چالش برانگیز است. ۱. تصویربرداری پزشکی اصولا فرایندی زمان بر و هزینه بر است و ۲. برچسب گذاری تصاویر پزشکی نیاز به افراد متخصص دارد و غالبا زمان بر است. برای حل این چالش میتوان از رویکرد های data augmentation استفاده کرده که در آن با تغییر روی داده های موجود به گونه ای که ویژگی های مهم آن محفوظ بماند، داده های جدیدی تولید میکنیم. تکنیک های کلاسیک data augmentation برای تصاویر شامل flipping و تغییر شدت پیکسل های تصویر، برش تصویر، rotation، نویز و بلور گاوسی، تغییرات رنگ و ... هستند. هرچند این تکنیک ها میتوانند در کاربرد های گوناگونی مفید باشند اما مستعد این هستند که باعث Overfit شدن مدل شوند خصوصا اگر دیتاست اولیه بیش از حد کوچک و نویزی باشد. به همین علت روش های data augmentation با استفاده از GAN مورد توجه قرار میگیرند. در این روش ها ما با استفاده از یک معماری GAN که روی حوزه ی تصاویر مورد بحث آموزش دیده است، تصاویر جدیدی در همان حوزه تولید میکنیم.

معماری GAN:



هم چنین استفاده از رویکردهای meta training و regularization برای بهبود عملکرد مدل هایی که روی داده های augmented آموزش میبینند میتواند در بهبود عملکرد آنها مفید باشد.

ج) شبکه های عصبی علی رغم داشتن الگوریتم های آموزش مشخص و قابل توضیح، مانند یک black box عمل میکنند و پس از آموزش چگونگی عملکرد آنها برای رسیدن به نتیجه ی نهایی شفاف نیست. هرچند معیارهای زیادی برای عملکرد این مدلها تعریف و استفاده میشود این معیارها اصولا چگونگی رسیدن مدل به پاسخ را توجه نمیکنند. هم چنین این معیار ها میتوانند به گونه های مختلفی تفسیر شوند و افراد با تغییرات جزئی در دیتا یا ... برداشت های متفاوتی از آن ها ارائه دهند که باعث میشود برای مسائل حساسی مانند تصاویر پزشکی قابل اعتماد نباشند. معیار های توضیح پذیری visual ضمن اینکه تفسیری از چگونگی عملکرد مدل برای رسیدن به جواب به ما ارائه میدهند برای کاربر قابل تفسیرتر نیز هستند و از همین رو میتوانند مفید باشند. LIME و CAM و grad-CAM مثالهایی از این دسته هستند.

د) روش های Model agnostic به روش هایی گفته میشود که وابسته به مدل نیستند و میتوانند برای توضیح پذیری طیف وسیعی از مدل ها استفاده شوند، این معیار ها به دلیل انعطاف پذیری بالا و قابلیت استفاده روی انواع مدل ها میتوانند به عنوان استاندارد برای مقایسه ی ساده و قابل درک مدل های مختلف استفاده شود. اما دقیقا به علت جنرال

بودن، نمیتوانند برای توضیح پذیری ویژگی های خاص هر مدل مفید باشند. روش های Model specific روش هایی هستند که ویژه ی دسته ی خاصی از مدل ها طراحی میشوند مانند GradCAM که تنها برای مدل های مبتنی بر داده های visual مانند CNN ها کاربرد دارد. مشخص است این مدل ها توانایی تبدیل شدن به یک استاندارد همگانی را ندارند ولی در عوض میتوانند در توضیح پذیری با جزئیات دقیق ویژگی های خاص مدل های حوزه ی خود مفید واقع شوند.