

سوال ①

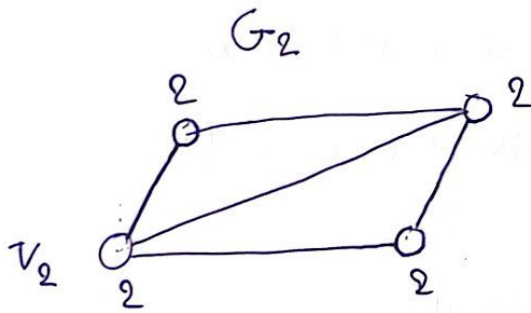
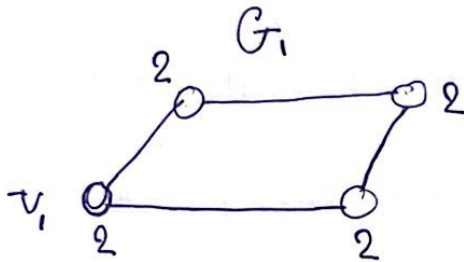
(۱) انتخاب تابع تجميع

دو گراف مقابل را در نظر می گیریم :

برای همی نودهای گراف مقدار اولیه ۲ را در نظر می گیریم .

پس $h_{v_1}^{(0)} = h_{v_2}^{(0)}$ برقرار است .

حال توابع تجميع مختلف صورت سوال را برای دو گراف افعال می کنیم .



$$\text{AGGREGATE}_{\max} : \begin{cases} h_{v_1}^{(1)} = \max(1, 1) = 1 \\ h_{v_2}^{(1)} = \max(1, 1, 1) = 1 \end{cases} \Rightarrow h_{v_1}^{(1)} = h_{v_2}^{(1)}$$

$$\text{AGGREGATE}_{\text{mean}} : \begin{cases} h_{v_1}^{(1)} = \frac{1+1}{2} = 1 \\ h_{v_2}^{(1)} = \frac{1+1+1}{3} = 1 \end{cases} \Rightarrow h_{v_1}^{(1)} = h_{v_2}^{(1)}$$

$$\text{AGGREGATE}_{\text{sum}} : \begin{cases} h_{v_1}^{(1)} = 1+1 = 2 \\ h_{v_2}^{(1)} = 1+1+1 = 3 \end{cases} \Rightarrow h_{v_1}^{(1)} \neq h_{v_2}^{(1)}$$

خوب همانطور که مشخص است به ازای تابع تجميع max و mean ، مقدار آیدیت شدن نود v_1 و v_2 در دو گراف G_1 و G_2 برابر است ($h_{v_1}^{(1)} = h_{v_2}^{(1)}$) اما به ازای تابع Sum برابر نیست .

2. آزمون وایسفالیر-لهمن (WL)

فرض: $\text{readout}(h_v^{(k)}, \forall v \in V_1) \neq \text{readout}(h_v^{(k)}, \forall v \in V_2)$

حکم: آزمون WL تصمیم می‌گیرد که گراف‌ها ایزومورف نیستند.

برای اثبات این حکم از برهان خلف و تناقض استفاده می‌کنیم.

فرض می‌کنیم آزمون WL بعد از k تکرار نمی‌تواند بگوید دو گراف G_1 و G_2 ایزومورف نیستند. این یعنی در k تکرار همواره بر حسب نودهای گراف G_1 و G_2 برابر خواهد بود. (با توجه به الگوریتم WL)

به بیان ریاضی \leftarrow بر حسب نودهای گراف G_1 و G_2 یعنی $\{h_v^{(i)}\}$ و همچنین

همسایه‌های آن‌ها یعنی $\{ \{h_u^{(i)}, u \in N(v)\} \}$ در

تکرار i و $i+1$ به ازای $k-1$ و $0, 1, 2, \dots$ یکسان و برابر است. چون اگر بر حسب نودها در تکرار i برابر نباشد آن‌ها

آزمون WL مجموعه بر حسب‌های مختلفی برای گراف G_1 و G_2 در تکرار $i+1$ بدست می‌آورد که خلاف فرض است.

از آنجایی که در تکرار 0 یعنی $i=0$ ، نودها دارای ویژگی یکسان $\chi(v)$ هستند و آزمون WL برای دو گراف یکسان G_1 و G_2 دارای بر حسب‌های $h_v^{(i)} = h_u^{(i)}$ است، می‌توانیم یک GNN embedding یکسان نیز برای u و v در تکرار i ام داشته باشیم. پس $h_v^{(i)} = h_u^{(i)}$ خواهد بود.

اگر این رابطه تکرار n ام برقرار باشد آنگاه در تکرار $n+1$ نیز برای هر جفت نود u, v رابطه $h_v^{(j+1)} = h_u^{(j+1)}$ برقرار است پس داریم:

$$\{(h_v^{(j)}, \{h_m^{(j)} : m \in N(v)\})\} = \{(h_u^{(j)}, \{h_m^{(j)} : m \in N(u)\})\}$$

نتیجه:

$$\{(h_v^{(j)}, \{h_m^{(j)} : m \in N(v)\})\} = \{(h_u^{(j)}, \{h_m^{(j)} : m \in N(u)\})\}$$

چون توابع aggregate و combine یکسان استفاده می شود پس به ازای ورودی یکسان، GNN خروجی یکسانی می دهد پس رابطه $h_v^{(j+1)} = h_u^{(j+1)}$ برقرار است.

این موضوع یعنی می توان ϕ ای یافت که به ازای هر نود در گراف G_1 و G_2 $\phi(h_v^{(i)}) = h_v^{(i)}$ باشد. چون $h_v^{(i+1)} = h_u^{(i+1)}$ است پس می توان ϕ را به این تفاسیر نتیجه گرفت $\{h_v^{(i+1)}\}$ به ازای همه نودها برقرار بوده و تابع readout برای دو گراف G_1 و G_2 یکسان است.

$$\text{readout}(h_v^{(K)} : \forall v \in V_1) = \text{readout}(h_v^{(K)} : \forall v \in V_2) :$$

نتیجه فوق خلاف فرض اولیه ما است و به تناقض رسیدیم.

پس آزمون WL بعد از K تکرار می تواند بگوید که G_1 و G_2 اینزومورف نیستند.

مقاله Hierarchical Graph representation

۱- رویکرد جدید این مقاله نسبت به مقاله های قبلی در زمینه استفاده از نمایش های گرافیک سلسله

مدرستی، استفاده همزمان ورودی Fine-level و Coarse-level است.

در واقع شبکه HACT-Net بصورت multi-level entity-based تقارن

یا توپولوژیک را نمایش می دهد. یعنی یک گراف برای نمایش سلول (Cell Graph) و

یک گراف برای نمایش بافت (Tissue Graph) دارد.

همچنین با استفاده ماتریسی Assignment سلول ها را به بافت های نظیرشان

انجام می دهد و اطلاعات هر سلول به بافت نظیرش برای تشخیص نوع بافت داده

می شود و این اطلاعات inter-level نیز به تشخیص این رویکرد کمک می کند.

۲- عکس های ورودی عکس بافت H&E Stained است. در این رویکرد، ابتدا

Stain normalization انجام می شود تا تنوعی که در عکس های H&E بدلیل

تکنیک های پیرولیک های مختلف وجود دارد کاهش بیاید.

سپس برای ساخت گراف HACT سه مرحله وجود دارد.

اول باید گراف Cell ساخته شود. برای اینکار ابتدا باید Nuclei، هسته های

Segmentation موجود در عکس را segment می کنیم و مختصات هر هسته نقطه ی مرکزی آن خواهد بود.

سپس ویژگی های مورفولوژیک با استفاده patch های به اندازه $w \times h$ که حول

مراکز هسته ها قرار می گیرند و همچنین شبکه ResNet استخراج می شود. و در نهایت

با استفاده از KNN، هسته های که از یک threshold خاصی به هم نزدیک تر باشند،

همسایه هم هستند و با یک edge به هم متصل می شوند. این کار با فاصله Euclidean

انجام می شود.

دوم باید گراف بافت (Tissue) ساخته شود. برای اینکار باید region های بافت ها را

با انجام یک Semantic segmentation مشخص کنیم. سپس برای استخراج ویژگی های بافت

همانند مرحله قبل با استفاده از patch های به سائز $w \times h$ و یک شبکه CNN ای، این

کار را انجام می‌دهیم. همچنین با الگوریتم SLIC که از ویژگی‌های K-means استفاده می‌کند مراکز بافت‌هایی که همسایه هم هستند را با هم merge می‌کنیم.

سوم باید ماتریس سلسه مراتب سلول به بافت را بسازیم. برای این کار سلول‌هایی که عضو یک ناحیه بافتی هستند باید به آن ناحیه مرتبط شوند. بدین منظور از یک ماتریس A استفاده می‌شود و سلسله زیر برای همه سلول‌ها و بافت‌ها پر می‌شود.

اگر سلول i عضو ناحیه بافتی j باشد. $A_{CG \rightarrow TG} [i, j] = 1$

اگر اینطور نباشد. $A_{CG \rightarrow TG} [i, j] = 0$

$A_{CG \rightarrow TG}$ یک ماتریس به ساینه $|V_{CG}| \times |V_{TG}|$ است.

سوال (2): مقایسه SlideGraph⁺

1) SlideGraph⁺ از یک نمایش گراف برای هر یک از عکس‌های WSI استفاده می‌کند که نودهای گراف نمایشگر یک سلول در WSI هستند. در واقع این گراف نمایشی از هندسه ارتباط میان سلولی آن عکس را بیان می‌کند.

در این روش مرفا Score نهایی WSI شخص نمی‌شود بلکه نودهای active یا تأثیرگذار در Score نهایی نیز قابل نمایش است.

عملیات‌های روش SlideGraph⁺ در وقت AUC بیشتر (حدود 0.8 در تست سب)

قابلیت تفسیر پذیری بیشتر به دلیل پشتیبانی در سطح نود (گره)، قابلیت نمایش هندسه ارتباط میان سلولی در عکس و همچنین efficient بودن از نظر میسبایی نسبت به مدل‌های مبتنی بر patch است.

تفسیر پذیری بهتر در SlideGraph⁺ به دلیل نمایش میزان تأثیرگذاری نودها در سبب Score نهایی است. نودهای active بخش‌هایی از تصویرند که احتمالاً سرطان تشخیص داده‌اند. (با مستعد بنظر می‌رسند)

② رنگ آمیزی DAB میزان بیان پروتئین HER2 را نشان می دهد که تعیین کننده وضعیت HER2 می باشد.

در روش رگرسیون چگالی DAB از یک تست paired از تصاویر H&E و IHC استفاده شده است. ابتدا این تصاویر pair با هم register شده اند و سپس از تصاویر IHC، درصد رنگ آمیزی DAB را می سب می کند. (این درصد در تصاویر IHC قابل اندازه گیری و تعیین می باشد)

سپس با یک مدل رگرسیون چگالی DAB سعی می کند چگالی DAB را با گرفتن تصاویر H&E می سب کند. این مدل نیاز به تصاویر IHC برای می سب چگالی DAB را برطرف می کند.

چون Case های مثبت HER2 دارای چگالی DAB بالا و Case های منفی HER2 دارای چگالی DAB پایین هستند، پس در پس بینی وضعیت HER2، چگالی DAB بسیار مهم است.

پس از اینکه چگالی DAB با مدل رگرسیون می سب شد بعنوان یک ویژگی در سطح نود در گراف شبکه عصبی برای تعیین وضعیت HER2 استفاده می شود و در نهایت این روش را بهبود می دهد.

مقاله ViT-V-Net

- محدودیت اصلی مدل‌های براساس ConvNet این است که نمی‌توانند روابط فضایی با فاصله زیاد از هم را پیدا کنند و تشخیص بدهند. این موضوع بدلیل locality ذاتی عملگرهای Convolution است. Voxel morph نیز بدلیل استفاده از ConvNet این محدودیت را دارد.

ViT-V-Net از ترکیب ConvNet و Transformer برای register تصاویر حجمی بصورت Self-supervised استفاده می‌کند. ViT به ویژگی‌های High-level تصاویر moving و fixed اعمال می‌شود و بایک شبکه روابط با فاصله دور از هم را یاد می‌گیرد. همچنین از skip connection های نقاط بلند برای حفظ اطلاعات محلی استفاده می‌شود.

- معماری: در این روش ابتدا عکس‌ها به ویژگی‌های High-level encode می‌شوند (با استفاده از لایه‌های Conv و max-pool) سپس این High-level feature ها به بخش ViT داده می‌شوند. این ویژگی‌ها به N Vector (برای هر patch یک vector به سائز $p \times c$) تبدیل شده‌اند.

سپس patch ها به فضای D بعدی latent map می‌شوند با استفاده از Projection خطی قابل train و patch embedding را ای‌دی می‌کند.

embedding های مکانی نیز به patch ها اضافه می‌شود تا اطلاعات مکانی حفظ شوند. سپس این پچ‌ها به بخش encode ترنسفورم داده می‌شود که حاوی 12 لایه multi-head self attention است. سپس خروجی ViT را تغییر ابعاد داده و به بخش decode شبکه می‌دهد تا در نهایت عکس با استفاده از spatial transformer ای‌دی شود.

شبیهت این روش با voxel morph در استفاده از لایه های Conv در بخش encode و decode است. همچنین ساختار U-Net از Voxel morph را حفظ کرده و دارای Skip Connection است.

از لحاظ استفاده از transformer برای پیدا کردن روابط فضایی دور، استفاده از مکانیزم multi-head attention، افزودن positional embedding به patch ها با روش Voxel morph متفاوت است.

- با اینکه روش V-Net میزان پذیرش و توجه شبکه به ارتباطات long-range را تا حدی بهتر کرد (نسبت به Voxel morph که دارای محدودیت در شناسایی رابطه های long-range واقع بود) اما مشکلاتی باقی مانده است:
- 1- زمینه های پذیرنده (receptive field) در چندین لایه اول بدلیل استفاده از کانولوشن و یک ساینز کرنل، محدود است و ویژگی های عمیقی (کلی تر) در لایه های عمیق تر شبکه دیده می شود.
 - 2- نشان داده شد که با محقق تر کردن لایه های Convolution، تأثیر Voxel های دور کمتر می شود.

روش حل این روش مبتنی بر این موضوع است که برعکس عملکرد کانولوشن که receptive field محدودی دارد، مکانیزم self-attention در ترنسفورمرها دارای لایه های receptive field مؤثر به اندازه ی بزرگ است. پس در این روش استفاده بسته از Swin transformer تا حد خوبی مشکلات مربوط به محدودیت Convolution را در Voxel morph و V-Net حل می کند.

- معماری: یک شبکه ترکیبی از transformer و ConvNet است که در آن Swin transformer به عنوان یک encoder برای درست آوردن تطابق رابطه فضایی بین عکس moving و عکس fixed استفاده می کند. همچنین از ConvNet برای بخش decode استفاده شود تا field جابجایی خروجی را ~~تولید کند~~ با استفاده از اطلاعات transformer تولید کند. همچنین مانند شبکه V-Net از skip connection های طولانی برای حفظ اطلاعات محلی بین encoder و decoder استفاده می کند.

نسبت به این روش با Voxel morph استفاده از skip connection و لایه های Conv در بخش decoder است. اما این روش از Swin transformer برای encode کردن استفاده می کند که با روش Voxel morph که مبتنی بر U-Net است متفاوت می باشد.