تمرین شمارهٔ ۳

مقدمه و چشم‌انداز: از ادراک در جراحی رباتیک تا چالش‌های بالینی دنیای واقعی

در دنیای پیشرفتهٔ جراحی رباتیک و سیستم‌های دستیار جراح خودمختار، ماژول «ادراک بصری» (Visual Perception) و به‌طور خاص تشخیص و ردیابی دقیق ابزارهای جراحی، ستون فقرات این سیستم‌ها را تشکیل می‌دهد. همان‌طور که فصل CNN و نوتبوک آموزشی آن بحث کردیم، شبکه‌های عصبی کانولوشنی (CNN) قلب تپندهٔ سیستم‌های ادراکی برای کارهایی نظیر ناوبری، درک صحنه و Grasping هستند. هرچند افق نهایی ما مداخله‌های کاملاً خودمختار رباتیک است، اما برای رسیدن به آن، سیستم‌های هوشمند ابتدا باید توانایی درک و پردازش جریان‌کاری در جراحی‌های پیچیدهٔ انسانی را به اثبات برسانند.

از این رو، در این تمرین از محیط‌های کنترل‌شدهٔ رباتیک فاصله گرفته و مستقیماً با تصاویر استخراج‌شده از ویدیوهای جراحی بالینی آب‌مروارید روبه‌رو خواهیم شد. شما باید مفاهیم تدریس شده در نوتبوک جامع CNN، از بلوک‌های پایهٔ کانولوشن و استراتژی‌های پیش‌پردازش گرفته تا تنظیم فراپارامترها با Optuna، یادگیری انتقال (Transfer Learning) و ابزارهای تفسیرپذیری نظیر Grad-CAM، را روی این دادهٔ واقعی پیاده‌سازی کنید تا یک ماژول ادراک بصریِ مقاوم (Robust) توسعه دهید که در آینده بتواند در یک لوپ کنترلی رباتیک مورد استفاده قرار گیرد.

۱ پرسش اول: شناخت دامنه، تحلیل اکتشافی داده‌ها (EDA) و استراتژی‌های پیش‌پردازش

این پروژه روی طبقه‌بندی ابزارهای جراحی در صحنه‌های عمل آب‌مروارید متمرکز است. با مراجعه به این لینک به دیتاست مربوطه دسترسی پیدا خواهید کرد. با توجه به مفاهیم بخش‌های ۶ تا ۸ نوتبوک آموزشی و مقالات مرجع، به موارد زیر پاسخ دهید:

آ. **درک سناریوی بالینی:** فرآیند جراحی آب‌مروارید را به‌دقت شرح دهید. فازهای اصلی این جراحی کدامند و در هر فاز چه ابزارهایی وارد میدان دید می‌شوند؟ چرا آگاهی بلادرنگ از فاز جراحی و ابزار موجود، برای یک بازوی رباتیک همکار یا سیستم دستیار هوشمند، حیاتی است؟

ب. **فرمول‌بندی مسئله:** تفاوت بنیادین طبقه‌بندی Single-label و Multi-label چیست؟ با توجه به اینکه در صحنهٔ جراحی آب‌مروارید ممکن است همزمان چندین ابزار (مثلاً پروب فاکو و چاپر) حضور داشته باشند، کدام رویکرد برای این پروژه الزامی است؟ در صورت انتخاب رویکرد Multi-label، با ارجاع به بخش ۹ نوتبوک، تابع زیان (Loss Function) مناسب چه خواهد بود و چرا دیگر نمی‌توان از Cross-Entropy استاندارد استفاده کرد؟

ج. **تحلیل داده‌ها و چالش‌های بصری:** در این دیتاست، داده‌ها به سه دستهٔ آموزش (Train)، اعتبارسنجی (Validation) و آزمون (Test) تقسیم شده‌اند (با یادآوری مبحث "چرایی و چگونگی" از بخش ۶ نوتبوک):

- تحلیل آماری و توصیفی دقیقی از توزیع کلاس‌ها ارائه دهید و نمودارهای مناسبی رسم کنید.
- تحلیل کنید که آیا تقسیم‌بندی فعلی داده‌ها توزیع کلاس‌ها را به‌خوبی حفظ کرده است یا با پدیدهٔ عدم تعادل کلاس‌ها (Class Imbalance) مواجهیم؟
- با توجه به چالش‌های تصاویر بالینی (مانند بازتاب نور میکروسکوپ، تارشدگی حرکتی ابزارها، و انسداد یا Occlusion)، کدام‌یک از روش‌های افزونگی داده (Data Augmentation) معرفی شده در بخش ۸

نوتبوک (نظیر Mixup، CutMix یا RandAugment) را برای مقاوم‌سازی مدل پیشنهاد می‌دهید؟ دلیل خود را با جزئیات توجیه کنید. اگر روش‌های دیگر خارج از نوتبوک هم می‌شناسید پیشنهاد کنید.

۲ پرسش دوم: طراحی معماری، تنظیم فرآیند آموزش و تفسیرپذیری مدل (XAI)

در این بخش باید قلب تپندهٔ سیستم ادراک خود را توسعه دهید. با الهام از آموزش‌های مربوط به معماری‌ها (بخش ۱۶ تا ۱۹) و جریان‌های کاری عملی (بخش ۲۰ تا ۲۲ نوتبوک)، موارد زیر را پیاده‌سازی و تحلیل کنید:

آ. **انتخاب معماری و یادگیری انتقال (Transfer Learning):** از بین معماری‌های مدرن (مانند ResNet، EfficientNet یا ConvNeXt)، مدلی را برای استخراج ویژگی‌ها و طبقه‌بندی ابزارها انتخاب کنید. چرا آموزش شبکه از صفر (From-scratch) در داده‌های پزشکی با حجم محدود، معمولاً به شکست می‌انجامد؟ نحوه استفاده از وزن‌های ازپیش‌آموزش‌دیده روی ImageNet را شرح دهید.

ب. **استراتژی آموزش و بهینه‌سازی:** با استفاده از فریم‌ورک Optuna (بخش ۱۴ نوتبوک)، فراپارامترهای کلیدی شبکه (شامل نرخ یادگیری، نوع بهینه‌ساز مانند AdamW در برابر SGD/Momentum، و Weight Decay) را جستجو و تنظیم کنید. برای مقابله با عدم تعادل کلاس‌ها، استفاده از کدام رویکرد در تابع زیان (مانند Focal Loss یا تکنیک Label Smoothing) را کارآمد می‌دانید؟ نمودارهای اتلاف (Loss) و دقت (Accuracy) را رسم کرده و وضعیت Overfitting یا Underfitting را به‌دقت تحلیل کنید.

ج. **ارزیابی جامع و تفسیرپذیری (Explainability):** با استفاده از دادهٔ آزمون (Test Data) عملکرد مدل را ارزیابی کنید. از آنجایی که متریک Accuracy در دیتاست‌های نامتعادل و مسائل Multi-label به‌شدت گمراه‌کننده است، نتایج را در قالب ماتریس درهم‌ریختگی (Confusion Matrix) و متریک‌های حیاتی دیگر نظیر Precision، Recall، F1-score و ناحیه زیر منحنی PR-AUC تحلیل کنید. در نهایت، برای اثبات اینکه مدل شما ابزارها را واقعاً بر اساس شکل هندسی خودشان یاد گرفته است و به نویزهای پس‌زمینهٔ جراحی تکیه نکرده، از تکنیک Grad-CAM (بخش ۲۲) برای تولید نقشه‌های حرارتی (Heatmaps) و یا از t-SNE (بخش ۲۱) برای بصری‌سازی فضای ویژگی‌های استخراج‌شده استفاده کنید و مشاهدات خود را تفسیر نمایید.

۳ پرسش سوم: اعتبارسنجی متقاطع، ارزیابی پایداری و یادگیری تجمعی

برای استقرار نهایی یک مدل در پردازشگر یک سیستم جراحی رباتیک، پایداری (Robustness) و قطعیت شبکه در مواجهه با شرایط پیش‌بینی‌نشده حرف اول را می‌زند. با تکیه بر بخش ۱۵ نوتبوک، سطح ارزیابی خود را ارتقا دهید:

آ. **رویکردهای اعتبارسنجی متقاطع (Cross-Validation):** مفاهیم K-Fold و Stratified K-Fold را مقایسه کنید. با توجه به توزیع فرکانس استفاده از ابزارها در فازهای مختلف جراحی آب‌مروارید، چرا استفاده از روش Stratified در این دامنه از اهمیت بالایی برخوردار است؟

ب. **پیاده‌سازی و ریفکتور کد:** پرسش دوم را با استفاده از رویکرد Stratified K-Fold مجدداً پیاده‌سازی کنید. توضیح دهید که برای این کار چه تغییراتی در ساختار داده‌ها، ساخت Dataloaderها (مثلاً استفاده از SubsetRandomSampler) و حلقهٔ آموزش پایه ایجاد کردید. واریانس و انحراف‌معیار عملکرد مدل در فولدهای مختلف را تحلیل کنید. آیا معماری شما نسبت به تغییرات داده‌های آموزش پایدار است؟

ج. **یادگیری تجمعی (Ensemble Learning) در برابر سرعت استنتاج:** حال که K مدل آموزش‌دیده از بخش قبل در اختیار دارید، چگونه می‌توانید از تکنیک‌هایی نظیر Ensemble Learning یا Model Soups (میانگین‌گیری پارامترهای مدل‌ها) جهت افزایش دقت نهایی استفاده کنید؟ با ارجاع به بخش ۲۳ نوتبوک (Inference Benchmarking)، بحث کنید که استفاده از روش Ensemble در فاز استنتاج (Inference) چه آسیبی به پردازش بلادرنگ (Real-time) در سیستم‌های رباتیک (نظیر افت Throughput یا افزایش Latency) وارد می‌کند و رویکرد Model Soups چگونه این تضاد را برطرف می‌سازد؟