



پردازش هوشمند تصاویر زیست پزشکی

نیم سال اول ۰۳-۰۲

مدرس: محمدحسین رهبان

مهلت ارسال: ۱۳ دی

تمرین پنجم

- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روز مشخص شده است.
- در طول ترم امکان ارسال با تاخیر تمرین‌ها بدون کسر نمره تا سقف ۱۲ روز وجود دارد. محل بارگزاری جواب تمرین‌ها بعد از ۴ روز بسته خواهد شد و پس از گذشت این مدت، پاسخ‌های ارسال شده پذیرفته نخواهند شد.
- توجه داشته باشید که نوت‌بوک‌های شما باید قابلیت بازاجرای ۱۰۰ درصد داشته باشند و در صورت نیاز به نصب یک کتابخانه یا دسترسی به یک فایل، مراحل نصب و دانلود (از یک محل عمومی) در نوت‌بوک وجود داشته باشد.
- هم‌فکری در انجام تمرین مانعی ندارد، فقط توجه داشته باشید که پاسخ تمرین حتما باید توسط خود شخص نوشته شده باشد. همچنین در صورت هم‌فکری در هر تمرین، در ابتدای جواب تمرین نام افرادی که با آن‌ها هم‌فکری کرده اید را حتما ذکر کنید.
- برای پاسخ به سوالات نظری در صورتی که از برگه خود عکس تهیه می‌کنید، حتما توجه داشته باشید که تصویر کاملاً واضح و خوانا باشد. در صورتی که خوانایی کافی را نداشته باشد، تصحیح نخواهد شد.
- محل بارگذاری سوالات نظری و عملی در هر تمرین مجزا خواهد بود. به منظور بارگذاری بایستی تمرین تئوری در یک فایل زیپ با نام `SPB_Theo_hw[HW-Number]_[First-Name]_[Last-Name]_[Student-Id].zip` و تمرین عملی نیز در یک فایل مجزای زیپ با نام `SPB_Prac_hw[HW-Number]_[First-Name]_[Last-Name]_[Student-Id].zip` بارگذاری شوند.
- در صورت وجود هرگونه ابهام یا مشکل، در کوئرای درس آن مشکل را بیان کنید و از پیغام دادن مستقیم به دستیاران آموزشی خودداری کنید.

بخش تئوری (۴۰ نمره)

۱. سوال اول (۱۰ نمره)

در مورد Object Detection به سوالات زیر پاسخ دهید.

- (i) نحوه‌ی عملکرد آشکارسازهای تک مرحله‌ای و دو مرحله‌ای را بیان کنید؟ تفاوت آن‌ها در چیست؟ موارد استفاده‌ی هر کدام در حوزه‌ی مدیکال چه مواردی است؟
- پاسخ:** در آشکارسازهای دو مرحله‌ای، ابتدا نواحی پیشنهادی وجود آبجکت توسط الگوریتم‌هایی مانند selective search مشخص می‌شود، سپس هر ناحیه به کلاس‌های موجود مسئله دسته‌بندی می‌شود. در آشکارسازهای تک مرحله‌ای، به جای خروجی دادن امتیاز confidence برای هر ناحیه، مستقیم امتیاز دسته‌بندی کلاس‌ها خروجی داده می‌شود.
- از آشکارسازهای تک مرحله‌ای در کاربرد های real time مانند اعمال جراحی استفاده می‌شود. از آشکارسازهای دو مرحله‌ای به دلیل دقت بالاتری که معمولاً دارند، در کاربرد های غیر برخطی که به دقت بالاتری نیاز دارند استفاده می‌شود.
- (ii) معماری R-CNN را شرح دهید. این آشکارساز چندمرحله‌ای عمل می‌کند؟ تفاوت آن با نسخه‌های fast R-CNN و faster R-CNN را بیان کنید؟

پاسخ: این شبکه توسط الگوریتم selective search حدود ۲۰۰۰ ناحیه پیشنهادی مشخص می کند. سپس هر ناحیه به وسیله ی اسکیل کردن یا RoI pool به ابعاد هم اندازه و استاندارد درمی آیند. در نهایت روی هر ناحیه پیشنهادی یک شبکه ی کانولوشنی اعمال می شود و یا سوار کردن هد های رگرسور و دسته بند، bounding box ها دقیق تر می شوند و کلاس آن ناحیه بدست می آید. در ورژن fast یکبار فروارد شبکه ی کانولوشنی طی می شود و از روی فیچر مپ بدست آمده، فیچر مپ های region of interest ها بدست می آیند. در نسخه ی faster تمامی ناحیه های پیشنهادی هم توسط یک شبکه ی عصبی بدست می آید. چراکه سرعت الگوریتم های سنتی بینایی کامپیوتر مانند selective search بسیار پایین است. به شبکه های پیشنهاد دهنده نواحی region proposal networks می گویند.

۲. سوال دوم (۱۰ نمره)

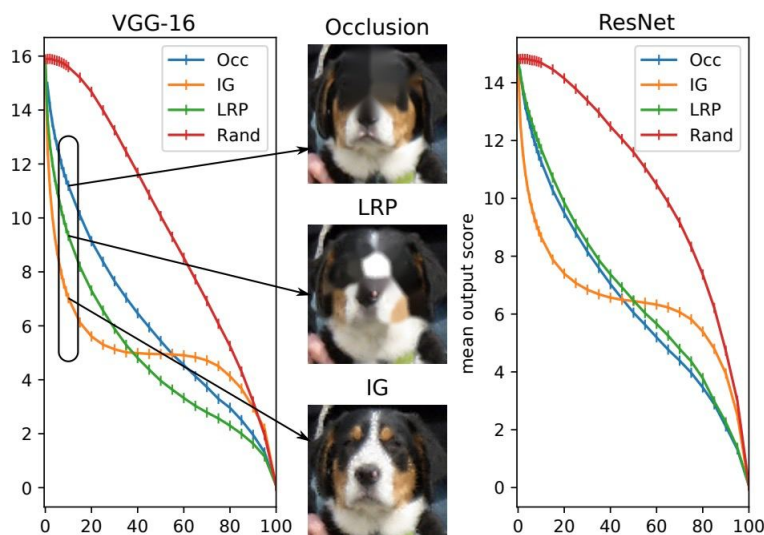
در مورد تفسیرپذیری به سوالات زیر پاسخ دهید

(i) در attribution methods خروجی آن یعنی saliency map ها دارای نویز هستند. برای وجود این نویز سه فرضیه مطرح هست. هر کدام را توضیح دهید و اگر برای آن راه حلی وجود دارد بیان کنید.

پاسخ:

فرضیه اول بیان میکند این مپ ها truthful هستند و واقعیت چیزی که شبکه به آن توجه می کند را بیان میکند. بنابراین لزومی ندارد که پیوسته و بدون نویز باشند. فرضیه دوم این نویز را به دلیل عدم پیوستگی در گرادیان به دلیل وجود اکتیویشن فانکشن هایی مثل ReLU بیان میکند. راه حل پیشنهادی با توجه به این فرضیه استفاده از روش های smoothing مثل gaussian moving average است. در فرضیه سوم، علت نویز را اشباع یا saturation اکتیویشن فانکشن هایی مثل ReLU یا Tanh بیان میکند. چرا که در محدوده اشباع، گرادیان این تابع ها به صفر میل میکند. در نتیجه برای حل آن از یک scaling با ضریب بین صفر و یک روی ورودی استفاده میکند تا مقادیر را حول مرکز و origin قرار دهد و از نواحی اشباع فاصله بگیرد.

(ii) در ارزیابی روش های attribution نحوه ی ارزیابی selectivity را توضیح دهید. برای توضیح این قسمت می توانید از الگوریتم ۳ این [survey](#) که در رابطه با pixel flipping است کمک بگیرید. همچنین به طور خاص توضیح دهید محورهای افقی و عمودی شکل زیر که در سروی آمده است نشان دهنده ی چه چیزی است؟



پاسخ: در این روش ها، قسمت هایی از تصویر که امتیاز saliency بالایی گرفته است را sort می کنند و به ترتیب شروع به مخدوش کردن یا حذف آن ها می کنند. چراکه میخواهند مشاهده کنند امتیاز confidence کلاس برنده در دسته بندی بعد از مخدوش کردن آن قسمت چقدر پایین می آید. در نتیجه در شکل، محور افقی درصد مخدوش کردن پیکسل هایی است که بر اساس امتیاز saliency یک روش تفسیر مرتب شده اند و محور عمودی کاهش امتیاز confidence را نشان می دهد.

۳. سوال سوم (۲۰ نمره)

در مورد Federated Learning به سوالات زیر پاسخ دهید.

(i) تفاوت کلیدی Federated Learning با یادگیری decentralized یا distributed learning چیست؟
پاسخ: در یادگیری decentralized داده ها به صورت i.i.d. روی سرور های مختلف پخش شده اند. اما در یادگیری فدرال، داده ها به صورت یکنواخت پخش نشده اند و از طرفی به هیچ وجه داده نمی تواند از سروری که روی آن قرار دارد خارج شود.

(ii) مقاله [Personalized Federated Learning with Theoretical Guarantees: A Model-Agnostic Meta-Learning Approach](#)

را مطالعه کنید. توضیح دهید در این مقاله چگونه از یک shared model به مدل های personalized برای هر کلاینت می رسد. علت ظهور مشتق مرتبه دوم (Hessian) را در الگوریتم این مقاله توضیح دهید.

Algorithm 1: The proposed Personalized FedAvg (Per-FedAvg) Algorithm

Input: Initial iterate w_0 , fraction of active users r .

for $k : 0$ to $K - 1$ **do**

Server chooses a subset of users \mathcal{A}_k uniformly at random and with size rn ;

Server sends w_k to all users in \mathcal{A}_k ;

for all $i \in \mathcal{A}_k$ **do**

Set $w_{k+1,0}^i = w_k$;

for $t : 1$ to τ **do**

Compute the stochastic gradient $\tilde{\nabla} f_i(w_{k+1,t-1}^i, \mathcal{D}_t^i)$ using dataset \mathcal{D}_t^i ;

Set $\tilde{w}_{k+1,t}^i = w_{k+1,t-1}^i - \alpha \tilde{\nabla} f_i(w_{k+1,t-1}^i, \mathcal{D}_t^i)$;

Set $w_{k+1,t}^i = w_{k+1,t-1}^i - \beta (I - \alpha \tilde{\nabla}^2 f_i(w_{k+1,t-1}^i, \mathcal{D}_t^i)) \tilde{\nabla} f_i(\tilde{w}_{k+1,t}^i, \mathcal{D}_t^i)$;

end for

Agent i sends $w_{k+1,\tau}^i$ back to server;

end for

Server updates its model by averaging over received models: $w_{k+1} = \frac{1}{rn} \sum_{i \in \mathcal{A}_k} w_{k+1,\tau}^i$;

end for

پاسخ: در قسمت سوم مقاله مرحله به مرحله به صورت کامل پاسخ سوال شرح داده شده است.