

به نام خدا



دانشگاه تهران



دانشکده مهندسی برق و کامپیوتر

**درس شبکه‌های عصبی و یادگیری عمیق**

**تمرین سوم**

نام دستیار طراح	نیما یغموری	پرسش ۱
رایانامه	nimayqmrnm@gmail.com	
نام دستیار طراح	فاطمه نائینیان	پرسش ۲
رایانامه	Fvma9596@gmail.com	
مهلت ارسال پاسخ	۱۴۰۲.۰۹.۱۷	

## فهرست

قوانین.....	۱
پرسش ۱. SAM.....	۱
۱-۱. آماده‌سازی مجموعه داده.....	۱
۲-۱. بارگذاری مدل.....	۱
۳-۱. تقویت داده.....	۱
(۵ نمره امتیازی).....	۲
۴-۱. بهینه‌ساز، متریک و تابع هزینه.....	۲
۵-۱. Fine-Tune کردن مدل.....	۲
۶-۱. ارزیابی نتایج.....	۳
پرسش ۲. آشنایی و پیاده‌سازی مدل Faster RCNN.....	۴
۲-۱. توضیحات مدل ها.....	۴
۲-۲. پیش پردازش.....	۴
۲-۳. آموزش شبکه.....	۵
۲-۴. بررسی داده های تست.....	۵

قبل از پاسخ دادن به پرسش‌ها، موارد زیر را با دقت مطالعه نمایید:

- از پاسخ‌های خود یک گزارش در قالبی که در صفحه‌ی درس در سامانه‌ی Elearn با نام **REPORTS\_TEMPLATE.docx** قرار داده شده تهیه نمایید.
- پیشنهاد می‌شود تمرین‌ها را در قالب گروه‌های دو نفره انجام دهید. (بیش از دو نفر مجاز نیست و تحویل تک نفره نیز نمره‌ی اضافی ندارد) توجه نمایید الزامی در یکسان ماندن اعضای گروه تا انتهای ترم وجود ندارد. (یعنی، می‌توانید تمرین اول را با شخص A و تمرین دوم را با شخص B و ... هم‌گروه باشید)
- **کیفیت گزارش شما در فرآیند تصحیح از اهمیت ویژه‌ای برخوردار است؛** بنابراین، لطفاً تمامی نکات و فرض‌هایی را که در پیاده‌سازی‌ها و محاسبات خود در نظر می‌گیرید در گزارش ذکر کنید.
- در گزارش خود مطابق با آنچه در قالب نمونه قرار داده شده، برای شکل‌ها زیرنویس و برای جدول‌ها بالانویس در نظر بگیرید.
- الزامی به ارائه توضیح جزئیات کد در گزارش نیست، اما باید نتایج بدست آمده از آن را گزارش و تحلیل کنید.
- **تحلیل نتایج الزامی می‌باشد، حتی اگر در صورت پرسش اشاره‌ای به آن نشده باشد.**
- **دستیاران آموزشی ملزم به اجرا کردن کدهای شما نیستند؛** بنابراین، هرگونه نتیجه و یا تحلیلی که در صورت پرسش از شما خواسته شده را به طور واضح و کامل در گزارش بیاورید. در صورت عدم رعایت این مورد، بدیهی است که از نمره تمرین کسر می‌شود.
- **کدها حتماً باید در قالب نوت‌بوک با پسوند ipynb تهیه شوند، در پایان کار، تمامی کد اجرا شود و خروجی هر سلول حتماً در این فایل ارسالی شما ذخیره شده باشد.** بنابراین برای مثال اگر خروجی سلولی یک نمودار است که در گزارش آورده‌اید، این نمودار باید هم در گزارش هم در نوت‌بوک کدها وجود داشته باشد.
- **در صورت مشاهده‌ی تقلب امتیاز تمامی افراد شرکت‌کننده در آن، 100- لحاظ می‌شود.**
- تنها زبان برنامه نویسی مجاز **Python** است.
- استفاده از کدهای آماده برای تمرین‌ها به هیچ وجه مجاز نیست.

- نحوه محاسبه تاخیر به این شکل است: پس از پایان رسیدن مهلت ارسال گزارش، حداکثر تا یک هفته امکان ارسال با تاخیر (به ازای هر روز یک درصد تجمعی) وجود دارد، پس از این یک هفته نمره آن تکلیف برای شما صفر خواهد شد.
- لطفا گزارش، کدها و سایر ضمایم را به در یک پوشه با نام زیر قرار داده و آن را فشرده سازید، سپس در سامانه‌ی Elearn بارگذاری نمایید:

HW[Number]\_[Lastname]\_[StudentNumber]\_[Lastname]\_[StudentNumber].zip

(مثال: HW1\_Ahmadi\_810199101\_Bagheri\_810199102.zip)

- برای گروه‌های دو نفره، بارگذاری تمرین از جانب یکی از اعضا کافی است ولی پیشنهاد می‌شود هر دو نفر بارگذاری نمایند.

در این پرسش، شما روی Fine-Tune کردن مدل جدید منتشر شده شرکت Meta برای مسئله Semantic Segmentation کار خواهید کرد. تمرکز اصلی، درک مفاهیم Segmentation تصویر، کار با مدل معرفی شده و تنظیم آن برای مسئله مورد نظر است.

### ۱-۱. آماده‌سازی مجموعه داده

(۵ نمره)

مجموعه داده انتخاب شده برای این پرسش، [تصاویر ماهواره‌ای از پهنه‌های آبی](#) می‌باشد. آن را دانلود کرده، تصاویر را خوانده و نمونه‌ای از آن‌ها را (تصویر اصلی در کنار ماسک) نمایش دهید. سپس داده‌ها را به نسبت ۱ به ۹ به دو بخش آموزش و ارزیابی تقسیم کنید.

### ۲-۱. بارگذاری مدل

(۱۵ نمره)

برای بارگذاری مدل از [مدل کارت](#) استفاده خواهید کرد. با توجه به [مقاله](#)، در مورد نحوه کار مدل توضیح داده و ساختار معماری مدل انتخاب شده را شرح دهید.

### ۳-۱. تقویت داده<sup>۲</sup>

(۱۵ نمره)

روش‌های تقویت داده، تغییراتی را در تصاویر از طریق انجام تکنیک‌هایی ارائه می‌کنند تا بدین ترتیب، توانایی مدل را در تعمیم‌دهی و Segmentation بهتر تصاویر در شرایط مختلف افزایش دهند. با توجه به ماهیت داده‌ها، روش‌های مناسبی را برای تقویت انتخاب کرده، علت انتخاب و نحوه تاثیرگذاری هر یک را توضیح داده و بعد از پیاده‌سازی، چند نمونه تقویت شده را نمایش دهید.

---

<sup>۱</sup> Segment Anything Model  
<sup>۲</sup> Data Augmentation

(۵ نمره امتیازی) کتابخانه [Albumentations](#) یک ابزار قدرتمند در زمینه تقویت داده‌ها برای مسائل بینایی ماشین و پردازش تصویر است. در صورت استفاده از توابع این کتابخانه برای پیاده‌سازی این بخش، ۵ نمره امتیازی به شما تعلق خواهد گرفت.

(پیشنهاد) توضیحات بخش پنجم این پرسش را قبل از انجام این بخش مطالعه کنید.

## ۴-۱. بهینه‌ساز، متریک و تابع هزینه


(۱۵ نمره)

دو متریک  $\text{IoU}^1$  Score و Dice Coefficient را پیاده‌سازی کرده و از آن‌ها به عنوان متریک در حین آموزش شبکه استفاده کنید. انتخاب بهینه‌ساز و تابع هزینه مناسب به اختیار شماست. در مورد تابع هزینه، می‌توانید از توابع آماده در اینترنت استفاده کنید اما لازم است توضیحات کافی در مورد علت انتخاب خود، کارکرد و چگونگی تنظیم آن ارائه دهید.

## ۵-۱. Fine-Tune کردن مدل

(۲۵ نمره)

جهت حفظ ویژگی‌های غنی از پیش آموزش دیده شده، دو جزء Vision Encoder و Prompt Encoder را Freeze کرده و مدل را تنها روی بخش Mask Decoder تنظیم کنید. همچنین بر اساس ماسک‌های موجود، Bounding Boxهایی طراحی کرده و به عنوان ورودی Prompt Encoder از آن‌ها استفاده کنید. از دادگان آموزشی برای آموزش مدل و از دادگان ارزیابی جهت بررسی کیفیت آموزش مدل (با گزارش مقادیر تابع هزینه و متریک) استفاده کنید.

(توجه) دقت شود ممکن است Bounding Boxها لزوماً در تمامی نمونه‌ها (به جهت گستردگی ماسک مربوطه در طول تصویر) منجر به راهنمای  برای مدل نشود؛ اما در مجموع می‌تواند تمرکز مدل را از نواحی اشتباه به سمت نواحی هدف بهبود بخشد. اهمیت انتخاب روش‌های صحیح در بخش سوم این پرسش، تاثیر خود را در این موضوع نشان خواهد داد.

---

<sup>1</sup> Intersection over Union

## ۱-۶. ارزیابی نتایج

(۲۵ نمره)

بعد از آموزش مدل به میزان کافی، ۵ نمونه از تصاویر ارزیابی را به همراه ماسک واقعی، ماسک پیش‌بینی شده و توزیع احتمال بدست آمده روی پیکسل‌های تصویر نشان دهید. تحلیل خود را از نتایج بدست آمده روی تصاویر و مقادیر محاسبه شده تابع هزینه و متریک‌ها در طی آموزش روی دادگان آموزشی و ارزیابی بیان کنید.

## پریش ۲. آشنایی و پیاده سازی مدل Faster RCNN

در این بخش می‌خواهیم تا یک مدل Faster RCNN را با یک دیتاست آموزش دهیم و نتایج را بررسی کنیم. برای این سوال مقاله Analysis of Object Detection Performance Based on Faster RCNN را شبیه سازی می‌کنیم. در مقاله برای آموزش از دیتاست Pascal VOC استفاده شده است. این دیتاست ۲۰ نوع کلاس و بیش از ده هزار عکس و annotation دارد. بنابراین با توجه به محدودیت های موجود، می‌خواهیم نتایج این شبیه سازی را برای یک دیتاست کوچکتر بررسی کنیم. می‌توانید به دیتاست مورد استفاده در این مساله با این [لینک](#) دسترسی پیدا کنید. در صورتی که تمایل به استفاده از پارامترهای متفاوتی از موارد گفته شده در مقاله دارد، لطفاً توجه خود را در گزارش ذکر کنید.

### ۲-۱. توضیحات مدل ها

(۲۰ نمره)

در رابطه با سه مدل RCNN، Fast RCNN و Faster RCNN تحقیق کنید و به صورت خلاصه بهبودهای مدل Faster RCNN را نسبت به دو مدل دیگر ذکر کنید. مقاله را مطالعه کنید و کاربرد هر بخش شبکه از جمله Conv Layer، RPN، ROI Pooling و Classification را توضیح دهید.

### ۲-۲. پیش پردازش

(۲۰ نمره)

تصاویر و برچسب های آن را از لینک داده شده استخراج کنید. ۵ نمونه از تصاویر Train را همراه با برچسب آن نمایش دهید. سپس باید ابعاد تصاویر را یکسان کنید. توجه فرمایید که هنگام resize ابعاد برچسب را نیز متناسب با تصاویر تغییر دهید.



## ۳-۲. آموزش شبکه

(۴۰ نمره)

مطابق اطلاعات مقاله، بخش های مختلف شبکه از جمله RPN، ROI Pooling و... را پیاده سازی کنید. پارامتر های داده نشده در مقاله را با بررسی خود انتخاب کنید و دلیل انتخاب خود را توضیح دهید. شبکه را با داده های Train و Validation آموزش دهید. نمودار loss را نیز نمایش دهید.

## ۴-۲. بررسی داده های تست

(۲۰ نمره)

حال ۵ نمونه از تصاویر segment شده داده های تست را همراه با bounding box واقعی و پیش بینی شده نمایش دهید. تحلیل خود را در مورد نتایج به دست آمده بیان کنید. نظر شما در مورد دقت های به دست آمده چیست و چگونه میتوان بهبودی در دقت داشت؟ بنظر شما ضعفی در مدل ها وجود دارد که بتوان با برطرف کردن آنها به بهبود دقت کمک کرد؟