

به نام خدا



دانشگاه تهران  
پردیس دانشکده‌های فنی  
دانشکده برق و کامپیوتر



## درس یادگیری عمیق با کاربرد در بینایی ماشین و پردازش صوت

تمرین شماره ۲

فروردین ۱۴۰۱

## فهرست

3.....مقدمه

4.....سوال 1

4.....پیش بینی نقشه برجستگی

4.....بخش الف

5.....بخش ب

5.....بخش ج

5.....بخش ه (امتیازی، ۱۵ درصد بارم سوال)

6.....سوال ۲

6.....بخش الف

7.....بخش ب

7.....بخش ج

7.....بخش ه

هدف از این تمرین آشنایی بیشتر شما با شبکه های عصبی کانولوشنی<sup>۱</sup> است. این دسته از شبکه ها در مسائل بسیاری کاربرد دارند که به دلیل برخورداری از ویژگی های اتصال محلی<sup>۲</sup> فیلتر ها و اشتراک گذاری پارامتر ها<sup>۳</sup> برای کار در زمینه تصاویر و ویدیو ها بسیار مناسب هستند و به نتایج بسیار خوبی دست پیدا کرده اند و پیشرفت های زیادی در این زمینه صورت گرفته است. به علاوه این شبکه ها در مسائل کار با متن و سیگنال های مختلف نیز کارآمد هستند. با این وجود کار با تصاویر به دلیلی که پیشتر ذکر شد محوریت اصلی این تمرین است.

---

<sup>1</sup> Convolutional Neural Networks (CNNs)

<sup>2</sup> local connectivity

<sup>3</sup> parameter sharing

## پیش بینی نقشه برجستگی



در بینایی کامپیوتر، پیش بینی نقشه برجستگی<sup>۱</sup> به معنای پیش بینی نواحی ای از تصویر است که بیشتر توجه یک فرد را در هنگام مشاهده جلب میکند و غالب تمرکز چشم و حواس اون بر این نواحی یا نقاط است. در واقع نقشه برجستگی میزان اهمیت پیکسل های مختلف یک تصویر را در سیستم بینایی انسان نشان میدهد. حل این مسئله کاربرد های بسیاری دارد. یک کاربرد بزرگ آن در فشرده سازی یا بریدن تصاویر و ویدیو ها است زمانیکه میخواهیم تشخیص دهیم کدام بخش از یک تصویر یا کدام نواحی از یک فریم ویدیو اهمیت بیشتری دارد که بتوان با حذف یا بریدن نواحی دیگر تصویر یا ویدیو را فشرده کرد به نوعی که کمترین میزان اطلاعات مورد نیاز برای درک تصویر از دست برود. کاربرد دیگر آن برای مثال کمک به تشخیص و شناسایی اشیا یا انسان ها در تصاویر است به نوعی که با توجه بیشتر به نواحی مهم محاسبات کمتری برای تشخیص شی لازم بوده و دقت افزایش میابد. همچنین در طراحی یک سایت، طراحی صحنه در فیلم ها و یا جانمایی اشیا، تشخیص نقشه برجستگی میتواند تاثیر زیادی در نشان دادن مقصود به بیننده یا کاربر داشته باشد. هدف این سوال پیاده سازی یک مدل ساده برای پیش بینی نقشه برجستگی است.

## بخش الف

در این بخش شما لازم تا مدل Deep Convnet از [این مقاله](#) را پیاده سازی کنید (پیشنهاد میشود برای درک بهتر مسئله مقاله را یک بار مطالعه کنید). شماتیک مدل و جزئیات ساختار آن در شکل ۴ از مقاله آورده شده و نکات بیشتر در بخش

<sup>1</sup> Saliency Map Prediction

۴.۱ توصیف شده است. برای راحتی کار و کاهش حجم محاسبات و حافظه مورد نیاز عمق شبکه های کانولوشنی ساختار یاد شده را به نصف کاهش دهید. برای مثال لایه کانولوشنی با عمق ۲۵۶ را به ۱۲۸ تقلیل دهید. دقت کنید که در سائز فیلتر ها تغییری ایجاد نکنید. برای آموزش مدل از دیتاست cat2000 استفاده کنید که دیتاست کوچکتري از دیتاست SALICON مقاله است. برای صرفه جویی در مصرف اینترنت این دیتاست را میتوانید با دستور زیر در colab دانلود کنید.

wget <http://saliency.mit.edu/trainSet.zip>

تمام جزئیات پیاده سازی را مشابه با آنچه در بخش ۴.۲ از مقاله آورده شده است پیش ببرید و از ۸۵ درصد داده ها برای آموزش<sup>۱</sup> و از باقی آنها برای ارزیابی<sup>۲</sup> مدل در طول آموزش استفاده کنید. دقت کنید که از تمام کلاس های دیتاست به صورت یکنواخت سمپل بگیرید (هم برای دادگان آموزش و هم ارزیابی). با توجه به تفاوت دیتاست ها اگر لازم دانستید هاپیر پارامتر های<sup>۳</sup> آموزش مدل را تغییر دهید. همانطور که در مسئله نیز ذکر شده است مدل خود را با تابع هزینه L2 آموزش دهید.

## بخش ب

نمودار هزینه بر حسب تکرار را برای دادگان آموزش و ارزیابی روی مدل خود رسم کنید و نحوه عملکرد آنرا بررسی کنید.

## بخش ج

برای شش تصویر از دیتاست، خود تصویر را به همراه نقشه برجستگی حقیقی و نقشه پیش بینی شده آن توسط مدل رسم کنید.

## بخش ه (امتیازی، ۱۵ درصد بارم سوال)

در رابطه با معیار ارزیابی AUC Judd مطالعه کرده و نحوه محاسبه آنرا شرح دهید. سپس این معیار را روی دادگان تست که با دستور زیر قابل دانلود هستند برای مدل خود محاسبه کنید و مختصری در رابطه با عملکرد مدل بر این اساس توضیح دهید.

wget <http://saliency.mit.edu/testSet.zip>

---

<sup>1</sup> train

<sup>2</sup> validation

<sup>3</sup> hyperparameters

## بازسازی تصویر از نقشه های ویژگی

میدانیم که در یادگیری ماشین، در هر مسئله بازنمایی هایی از داده ها مناسب است که ویژگی زائد از آنها حذف شده باشد در حالیکه اطلاعات و ویژگی های مفید برای استنتاج از داده ها حفظ شده باشد. شبکه های عصبی کانولوشنی در بسیاری از مسائل اثبات کرده اند که توانایی برآورده کردن این شرط در بازنمایی داده ها را دارند. در این مسئله هدف بررسی صحت این امر است. در این سوال با پیاده سازی یک مدل ساده متوجه خواهید شد که چه ویژگی هایی از یک تصویر توسط مدل کانولوشنی حذف و چه ویژگی هایی حفظ میشوند.

## بخش الف

در این بخش لازم است مدل هایی را برای بازسازی تصویر از نقشه های ویژگی استخراج شده از لایه های مختلف یک شبکه کانولوشنی آموزش دهید. مبنای این سوال [این مقاله](#) است. در این مقاله نویسندگان ساختار مدل AlexNet که بر روی دیتاست ImageNet برای مسئله طبقه بندی تصاویر آموزش دیده است را در نظر میگیرند که ساختار آن در جدول ۳ از مقاله آمده است. سپس با استفاده از بردار های ویژگی خروجی لایه های مختلف آن، قصد دارند تا تصاویر اولیه را بازسازی کنند.

در این تمرین به جای استفاده از دیتاست ImageNet از دیتاست Tiny ImageNet استفاده میکنید. از میان ۲۰۰ کلاس موجود در دیتاست ۲۰ کلاس با به صورت تصادفی انتخاب کنید و شناسه آنها را در گزارش خود بیاورید. از هر کلاس ۵۰ تصویر را به عنوان داده ارزیابی و باقی را برای آموزش مدل استفاده کنید. دقت کنید که مدل را با نمونه گیری یکنواخت از تمام کلاس ها آموزش دهید (هم برای آموزش هم ارزیابی). برای صرفه جویی در مصرف اینترنت از دستور زیر برای دانلود دیتاست در colab استفاده کنید.

wget http://cs231n.stanford.edu/tiny-imagenet-200.zip

شما لازم است تا از میان ۸ لایه مختلف AlexNet مدل هایی را برای بازسازی تصویر از خروجی چهار لایه conv2, conv5, fc6, fc8 آموزش دهید. برای خروجی لای های کانولوشنی از ساختار معرفی شده در جدول ۷ و برای خروجی لایه های خطی از ساختار معرفی شده در جدول ۸ بهره بگیرید (چهار مدل مجزا آموزش دهید). برای کاهش محاسبات و حافظه عمق شبکه ها کانولوشنی را به نصف تقلیل دهید.

دقت کنید که وزن های شبکه AlexNet را در زمان آموزش مدل های کدگشا<sup>۱</sup> ثابت در نظر بگیرید و آموزش ندهید. از مدل از پیش آموزش دیده AlexNet<sup>۲</sup> در پکیج torchvision استفاده کنید. جزئیات و هایپر پارامتر های آموزش مدل ها را

<sup>۱</sup> decoder

<sup>۲</sup> pretrained

از بخش ۲.۲ بگیرید و در صورت نیاز آنها را تغییر دهید. همانطور که در مقاله ذکر شده است از تابع هزینه L2 برای آموزش استفاده کنید.

## بخش ب

نمودار هزینه بر اساس تکرار را برای دادگان آموزش و ارزیابی در طول آموزش رسم کنید و آنها را بررسی کنید.

## بخش ج

۵ کلاس از ۲۰ کلاس موجود را انتخاب کنید و از هر کدام ۱ تصویر از دادگان آموزش و ۱ تصویر از دادگان ارزیابی را انتخاب کنید و خود تصویر ها را به همراه تصاویر بازسازی شده آنها، از خروجی هر چهار لایه را، رسم کنید.

## بخش ه

با دقت در تصاویر بخش ب، در رابطه با ویژگی هایی که توسط یک شبکه عصبی کانولوشنی در هر لایه حذف یا حفظ میشوند تحلیل خود را ارائه دهید. در تحلیل خود لایه های مختلف را هم با هم بررسی کنید و نتیجه بگیرید. میتوانید با مطالعه مقاله تحلیل های خود را قوی تر و علمی تر کنید.

## نکات:

- مهلت تحویل این تمرین، پنج شنبه 18 فروردین است.
- انجام این تمرین به صورت یک نفره می باشد.
- شما قادر نیستید هیچ تمرینی را با بیش از 7 روز تاخیر بارگذاری کنید (دقیقا 7 روز پس از مهلت آپلود، سامانه بسته خواهد شد).
- گزارش شما در فرآیند تصحیح از اهمیت ویژه ای برخوردار است. لطفاً تمامی نکات و فرض هایی که برای پیاده سازی ها و محاسبات خود در نظر می گیرید را در گزارش ذکر کنید. دقت داشته باشید 50 درصد از نمره تمرین شما مربوط به گزارش است.
- کدهای خود را به صورت عکس در داخل گزارش کپی نکنید و با فرمتی مناسب آن را در گزارش قرار دهید.
- داخل کدها کامنت های لازم را قرار دهید و تمامی موارد مورد نیاز برای اجرای صحیح کد را ارسال کنید.
- الزامی به ارائه توضیح جزئیات کد در گزارش نیست. اما باید نتایج بدست آمده را گزارش و تحلیل کنید.
- گزارش را در قالب تهیه شده که روی صفحه درس در سامانه eLearn بارگذاری شده، بنویسید. در صورت تمایل می توانید از Latex نیز برای نوشتن گزارش استفاده نمایید.
- در گزارش خود برای تصاویر زیرنویس و برای جداول هم بالانویس اضافه کنید.
- برای انجام این تمرین فقط مجاز به استفاده از زبان برنامه نویسی Python هستید و امکان استفاده از کتابخانه های = یادگیری عمیق نظیر Tensorflow و PyTorch را ندارید.
- از آدرس دهی مطلق در کدهای خود استفاده نکنید و به جای آن از آدرس دهی نسبی استفاده نمایید.
- فایل های ارسال شده باید به فرمت py. باشد و از ارسال فایل تمرین ها به صورت ipynb خودداری نمایید. همچنین ساختار کلی کدهای شما باید حداقل شامل فایل های زیر باشد. همچنین پیشنهاد می گردد قسمت های بارگذاری داده، توابع مورد استفاده در فایل هایی مانند dataloader و utils قرار داده شود.

نام فایل	توضیح
model	ساختار مدل
main	کد آموزش و اجرای مدل

- کد شما باید قابلیت اجرا بر روی قسمت کوچکی از داده ها را داشته باشد تا دستیار آموزشی مربوطه بتواند با استفاده از کد شما در مدت زمان کوتاهی مدل شما را آموزش دهد.
- در صورت مشاهده ی موارد تشابه بین دو یا چند فرد در گزارش کار و یا کد ، به طرفین تقلب نمره صفر داده خواهد شد. کپی برداری از کدهای آماده موجود در اینترنت و یا استفاده از کدهای افراد ترم های گذشته تفاوت چندانی با تقلب ندارد.



- اگر بخشی از کد را از کدهای آماده اینترنتی استفاده میکنید که جزء قسمتهای اصلی تمرین نمیباشد، حتما باید لینک آن در گزارش و کد ارجاع داده شود، در غیر اینصورت تقلب محسوب شده و کل نمره تمرین را از دست میدهید
- لطفا فایل کدها و سایر ضمائم مورد نیاز را با فرمت زیر در صفحه درس در سامانه eLearn بارگذاری نمائید.  
HW2\_[Lastname]\_[StudentNumber].zip
- در صورت وجود هرگونه ابهام یا مشکل می‌توانید از طریق رایانامه زیر با دستیار آموزشی طراح تمرین در تماس باشید :

**[p.baghershahi@ut.ac.ir](mailto:p.baghershahi@ut.ac.ir)**