به نام خدا

دانشگاه تهران

پردیس دانشکده‌های فنی

دانشکده مهندسی برق و کامپیوتر

درس یادگیری عمیق با کاربرد در بینایی ماشین و پردازش صوت

تمرین شماره ۲

فروردین ۱۴۰۰

❖ مقدمه

همانطور که می‌دانید با روی کار آمدن شبکه‌های عمیق کانولوشنی، تحول عمیقی در حوزه یادگیری ماشین صورت گرفت. این شبکه‌ها در ساختارهای مختلف و کاربردهای مختلفی از جمله طبقه‌بندی، تشخیص اشیا، تقسیم‌بندی تصاویر، تشخیص موقعیت مفاصل و غیره کاربرد دارند. در این تمرین هدف آشنایی با بعضی از کاربردهای شبکه‌های کانولوشنی می‌باشد. از آنجایی که پیاده‌سازی این شبکه‌ها با پیچیدگی‌های زیادی همراه می‌باشد، بنابراین شما مجاز هستید که در این سری از تمرینات از امکانات کتابخانه Pytorch استفاده نمایید.

❖ سوالات

• سوال اول

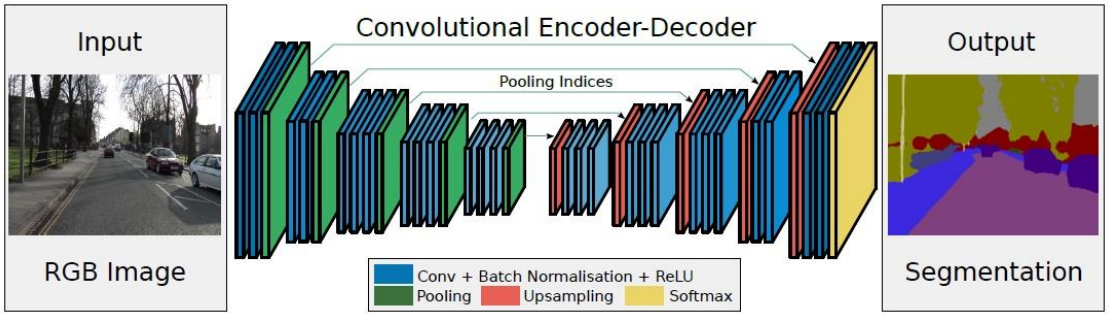
در سوال اول می‌خواهیم به موضوع تقسیم‌بندی معنایی تصاویر یا به عبارتی Semantic Segmentation بپردازیم. مدل‌های مختلفی در این حوزه وجود دارند که با شیوه‌های مختلف به دنبال بدست آوردن بیشترین دقت در تقسیم‌بندی تصاویر هستند. به عبارتی یک تصویر می‌تواند از بخش‌های مختلفی مانند آسمان، درخت، حیوان، انسان و غیره تشکیل شده باشد، هدف این است که به هر پیکسل در این تصویر یک برچسب که مربوط به کلاس آن پیکسل است تخصیص داده شود و در نهایت با رنگ‌بندی این پیکسل‌ها، کلاس‌ها یا بخش‌های مختلف تصویر را مشخص نماییم. در شکل شماره۱ نمونه‌ای از این تقسیم‌بندی را مشاهده می‌کنید.

شکل۱: نمونه‌ای از تقسیم‌بندی معنایی تصاویر که تصویر سمت چپ به بخش مختلف شامل درخت، آسمان، جاده و غیره تقسیم شده‌است

حال برای این هدف ما می‌خواهیم از مدل SegNet که مدلی بر پایه معماری Encoder-Decoder می‌باشد استفاده نماییم. همچنین دیتاست مورد استفاده در این تمرین Camvid می‌باشد که در ادامه درباره آن صحبت خواهیم کرد.

○ بخش اول

همانطور که پیش‌تر توضیح دادیم، مدل بکاررفته در این تمرین، مدل پایه Segnet می‌باشد که شکل کلی مدل اصلی Segnet را در تصویر شماره۲ مشاهده می‌کنید.



شکل۲: مدل Segnet جهت تقسیم‌بندی معنایی تصاویر

جزییات مربوط به این شبکه را می‌توانید در مقاله زیر مشاهده نمایید.

https://arxiv.org/abs/1511.00561

ابتدا مقاله فوق را مطالعه نمایید و سپس به صورت کامل نحوه عملکرد این شبکه را توضیح دهید.

از آنجایی که تعداد پارامترهای این شبکه زیاد می‌باشد و ممکن است نیاز به زمان زیادی برای آموزش داشته باشد، ما از مدل پایه این شبکه به نام Segnet-base که اطلاعات آن را در مقاله زیر می‌توانید مشاهده نمایید، استفاده می‌کنیم.

https://arxiv.org/abs/1505.07293

این مدل بر خلاف مدل اصلی از ۴ لایه انکودر و دیکودر استفاده کرده‌است و البته تفاوت‌های کوچکی با مدل اصلی دارد. با مطالعه این مقاله ساختار این شبکه را برای هر لایه ترجیحا در یک جدول در گزارش کار خود بیان کنید.

○ بخش دوم

در این مرحله شما نیاز دارید که با یکی از دیتاست‌های بکار رفته در این حوزه آشنا شوید. دیتاست استفاده شده در این قسمت به نام CamVid می‌باشد که شامل ۷۰۱ تصویر مربوط به محیط باز که از تصاویر ویدئویی استخراج شده است، می‌باشد. این دیتاست از سه فایل مهم تشکیل شده‌است که باید به صورت مجزا دانلود شود.

فایل تصاویر اصلی مربوط به این دیتاست را می‌توانید از لینک زیر دانلود نمایید.

https://s3.amazonaws.com/fast-ai-imagelocal/camvid.tgz

فایل مربوط به تصاویر برچسب خورده متناظر با هر تصویر اصلی را از لینک زیر دانلود نمایید.

http://mi.eng.cam.ac.uk/research/projects/VideoRec/CamVid/data/LabeledApproved_full.zip

فایل مربوط به کلاس‌های مختلف بکار رفته در این دیتاست را از لینک زیر دانلود نمایید.

http://mi.eng.cam.ac.uk/research/projects/VideoRec/CamVid/data/label_colors.txt

توجه داشته باشید که ما به پوشه label که از مسیر اول دانلود نموده‌اید، نیازی نداریم. حال با بررسی فایل‌های دانلود شده، به صورت مختصر در گزارش خود درباره نحوه استفاده از این دیتاست در مدل و عملیات پیش‌پردازشی که برای آن نیاز دارید، توضیح دهید. توجه نمایید که نیاز به تغییر ابعاد تصاویر اصلی وجود دارد.

- **بخش سوم**

در این قسمت مدل Segnet-base را به طور کامل پیاده‌سازی نمایید و با انتخاب هایپر پارامترهای مناسب و همچنین اندازه بچ مناسب، شبکه پیاده‌سازی شده را آموزش داده و نمودار خطا داده آموزش و تست را در طول یادگیری رسم کرده و در گزارش خود بیان نمایید. سپس در پایان آموزش سه تصویر مختلف را به شبکه آموزش داده شده بدهید و خروجی Segment شده آن را با رنگ‌های مناسب ترسیم کرده و با نسخه اصلی آن مقایسه نموده و در گزارش خود ارائه دهید.

- **بخش چهارم**

با اضافه نمودن لایه Batch Normalization بعد از هر لایه کانولوشنی در مدل بکار رفته در قسمت قبل، عملیات مربوط به بخش سوم را مجددا تکرار نمایید و نتایج را با قسمت قبل مقایسه نمایید.

- **سوال دوم**

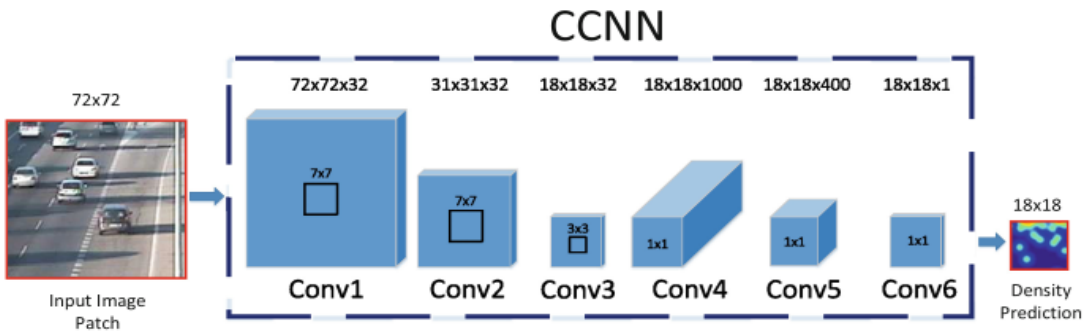
شکل ۳: نمونه ای از تصاویر پر ازدحام در مسأله شمارش اشیا [2]

- **مقدمه**

مسئله شمارش اشیا (Object Counting) در تصاویر و ویدیو ها، کاربرد های فراوانی در زمینه های گوناگون دارد. این مسئله خود در قالب های مختلف و برای شمارش موجودیت های متفاوتی از جمله شمارش در ازدحام (crowd counting)، سلول های میکروسکوپی، حیوانات، وسایل نقلیه، ساختمان ها، برگ ها و عوارض طبیعی تعریف میشود. در این تمرین قصد داریم تا یک نمونه از این مسائل را بررسی کنیم.

یکی از روش های پرطرفدار در حل مسئله شمارش اشیا، روش مبتنی بر نقشه چگالی (density map-based) است که در این تمرین از ان بهره خواهیم برد. انچه در ادامه توضیح داده خواهد شد برگرفته از کارهای صورت گرفته در [۱] و [۲] است که انتظار میرود برای اشنایی بهتر با موضوع و رفع ابهامات انها را مطالعه کنید.

مدلی که استفاده خواهیم کرد مشابه شبکه CCNN است که توسط [۱] معرفی شده و در شکل۴ قابل مشاهده است. این ساختار مبتنی بر شبکه های عصبی حلقوی (convolutional neural networks—CNN) است. این مدل در ابتدا تصویر هدف را به همراه مختصات دو بعدی متناظر ان که مربوط به اشیا یا موجودیت های مختلف است را به عنوان ورودی به یک کانولوشنی میدهد و در نهایت یک نقشه چگالی به عنوان خروجی تولید میکند که پیکسل های با مقدار بزرگتر ان بیانگر مختصات اشیا پیش بینی شده است. سپس این نقشه چگالی پیش بینی شده میبایست با نقشه چگالی حقیقی (ground truth) حاصل از مختصات نشانه گذاری شده اشیا مقایسه شود تا مقدار خطا تعیین شده و برای اموزش شبکه استفاده شود.



شکل ۴: ساختار شبکه CCNN معرفی شده در [1]

○ نقشه چگالی مختصات نقاط

برای تولید نقشه چگالی از مختصات حقیقی اشیا ابتدا پیکسل های مربوط به مختصات دقیق هر شی در یک تصویر صفر مقدار دهی میشوند، سپس با اعمال یک فیلتر گوسی بر آن تصویر نهایی مربوط به چگالی نقاط تولید میشود. دو نمونه از این تصاویر در شکل ۵ قابل مشاهده هستند. این نقشه های چگالی معیار سنجش خطای خروجی مدل CCNN هستند.

شکل ۵: نمونه هایی از نقشه چگالی

○ ساختار شبکه

همانطور که اشاره شد ساختار مورد استفاده مشابه با ساختار پیشنهادی CCNN با تفاوت هایی که در جدول ۱ مشخص شده است.

جدول ۱: ساختار مدل

Layer	Input channels	Output channels	Kernel size	Stride	Padding
Conv1	1	32	11*11	1	5
ReLU					
Conv2	32	32	7*7	1	3
Max_Pool1	-	-	2*2	2	0
ReLU					
Conv3	32	64	5*5	1	2
Max_Pool2	-	-	2*2	2	0
ReLU					

Conv4	64	1000	1*1	1	0
ReLU					
Conv5	1000	400	1*1	1	0
ReLU					
Conv6	400	1	1*1	1	0

○ دیتاست

در این تمرین از دیتاست Shanghaitech که توسط [۲] ارائه شده است استفاده خواهیم کرد. برای دسترسی به این دیتاست میتوانید از این لینک استفاده کنید. این دیتاست خود دارای دو بخش با عنوان part_A و part_B است که در هر بخش تصاویر به دو دسته train و test تقسیم شده اند. part_A این دیتاست دارای تصاویر با ازدحام جمعیت بالا و ابعاد متفاوت است. part_B دارای تصاویر ثبت شده از نواحی شهری با ازدحام کمتر است که این تصاویر دارای ابعاد ثابت ۷۱۳*۱۰۲۴ هستند. برای اسودگی کار ما تنها از بخش B یا part_B این دیتاست استفاده خواهیم کرد که دارای ۴۰۰ تصویر اموزشی و ۳۱۶ تصویر تست است. همچنین متناظر با هر تصویر مختصات دو بعدی سر افراد موجود در تصویر نیز در فایل جداگانه ای اورده شده است. برای توضیحات بیشتر میتوانید از منبع [۲] استفاده کنید.

○ پیاده سازی

۱- مدل توصیف شده در بخش ساختار شبکه را پیاده سازی کنید.

۲- از انجایی که تعداد تصاویر این دیتاست کم است، با توجه به راهکار شرح داده شده در [۲] از هر تصویر ۹ عدد patch با ابعاد ۲۶۰*۲۶۰ به عنوان سمپل و به صورت رندوم استخراج کنید. برای اموزش شبکه از این تصاویر (نه از تصاویر اصلی) استفاده کنید. دقت کنید که پس از استخراج هر تصویر باید نقاط نشانه گذاری شده متناظر با هر patch را نیز فیلتر کرده و برای سنجش خطا و ارزیابی مدل استخراج کنید. (راهنمایی: دقت کنید که پس از استخراج مختصات نقاط مربوط به هر patch باید دستگاه مختصات خود را نسب مختصات patch انتخاب شده جا به جا کنید و مختصات جدید نقاط را محاسبه کنید)

۳- همانطور که در توضیحات بخش نقشه چگالی توضیح داده شد، باید متناظر با هر دسته از مختصات نقاط نشانه گذاری شده برای هر تصویر یک نقشه چگالی نقاط ایجاد کنید. از انجایی که در ساختار شبکه از دو لایه max pooling استفاده شده است، ابعاد تصویر در انتها $\frac{1}{4}$ ابعاد اولیه خواهند بود. به این جهت پیش از تولید نقشه چگالی باید مختصات نقاط نشانه گذاری شده را down scale کنید. سپس از کرنل گوسی با standard deviation برابر با ۳ استفاده کنید که مرکز این کرنل گوسی همان نقطه نشانه گذاری شده است. برای تولید این کرنل ها میتوانید از پکیج ها یا کد های اماده نیز استفاده کنید.

۴- با مقایسه تصویر خروجی شبکه برای هر تصویر و نقشه چگالی حقیقی که برای ان ایجاد کرده اید میزان خطای رگرسیون اقلیدوسی را که در رابطه ۴ از [۱] امده است محاسبه کنید و از ان برای اموزش شبکه استفاده کنید.

○ سوالات

۱- مدل خود را مطابق بخش پیاده سازی تکمیل کرده و اموزش دهید. میزان loss و معیار های MAE و MSE را مطابق رابطه ۲ از [۲] را در طول اموزش برای داده های اموزش و اعتبار سنجی گزارش کنید. پس از انکه از هر تصویر patch ها را استخراج کردید از ۱۰ درصد داده ها برای validation استفاده کنید. پس از اموزش شبکه عملکرد انرا بر روی داده های تست بسنجید و معیار های MAE و MSE را برای دادگان تست نیز گزارش کنید.
۲- پس از تکمیل اموزش شبکه، برای ۴ patch از تصاویر به صورت نمونه، نقشه چگالی نقاط و تعداد افراد تشخیص داده شده در ان تصاویر را گزارش کنید.
۳- با استفاده از imgaug دیتاست خود را با تبدیل های مختلف از جمله rotation, translation, scaling در هر دو جهت x و y و یا تغییر در illumination و contrast به ۱,۲ برابر افزایش دهید و مجدد شبکه را با داده های جدید اموزش داده و نتایج را با حالت قبل مقایسه و تحلیل کنید. برای مقایسه از جدول و نمودار بهره بگیرید.

○ امتیازی

همانطور که در این تمرین با ساختار CCNN مشاهده کردید، در این شبکه از هیچ لایه fully connected استفاده نشده است. ساختار شبکه [3] که در شکل ۳ از این مرجع قابل مشاهده است، کاملا مشابه با ساختار CCNN است با این تفاوت که به جای استفاده از لایه عمیق کانولوشنی با اندازه کرنل ۱*۱ از لایه fully connected در ان استفاده شده است. نویسندگان مقاله CCNN ادعا میکنند که عملکرد شبکه پیشنهادی انها با تغییر لایه fully connected به لایه عمیق کانولوشنی بهبود یافته است. با تغییر لایه های Conv4, Conv5 و Conv6 به لایه fully connected و ایجاد تصویر چگالی نقاط مشابه روش [۳]، مدل خود را اموزش دهید و صحت این ادعا را بررسی کنید. همچنین در رابطه با نقاط قوت یا ضعف افزایش عمق شبکه در مقابل افزایش عرض آن توضیح دهید.

مـر ا جـع

[1] D. Oñoro-Rubio and R. J. López-Sastre, "Towards perspective-free object counting with deep learning," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9911 LNCS, pp. 615–629, 2016, doi: 10.1007/978-3-319-46478-7_38.

[2] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, "Single-image crowd counting via multi-column convolutional neural network," 2016, doi: 10.1109/CVPR.2016.70.

[3] C. Zhang, H. Li, X. Wang, and X. Yang, "Cross-scene crowd counting via deep convolutional neural networks," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June-2015, pp. 833–841, 2015, doi: 10.1109/CVPR.2015.7298684.

نکات:

- مهلت تحویل این تمرین، تا پایان روز شنبه ۴ اردیبهشت ماه می‌باشد.
- انجام این تمرین به صورت انفرادی می‌باشد.
- برای انجام این تمرین استفاده از امکانات کتابخانه Pytorch بلامانع می‌باشد.
- داخل کدها کامنت‌های لازم را قرار دهید و تمامی موارد مورد نیاز برای اجرای صحیح کد را ارسال کنید.
- **در صورت مشاهدهٔ موارد تشابه بین دو یا چند فرد در گزارش کار و یا کد ، به طرفین تقلب نمره صفر داده خواهد شد و هیچ گونه عذر و بهانه‌ای از جمله ارسال کد به دوست خود و عدم آگاهی از کپی برداری کد شما پذیرفته نخواهد شد، بنابراین به هیچ عنوان کدهای خود را در اختیار دیگران قرار ندهید در غیر این صورت مسئولیت تقلب بر عهده شما نیز می‌باشد. همچنین کپی برداری از کدهای آماده موجود در اینترنت و یا استفاده از کدهای افراد ترم‌های گذشته تفاوت چندانی با تقلب ندارد و در چنین مواردی نیز نمره صفر به فرد تعلق می‌گیرد و جای هیچگونه اعتراضی وجود 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.
- گزارش شما در فرآیند تصحیح از اهمیت ویژه‌ای برخوردار است و نیمی از نمره شما را دربرخواهد گرفت. لطفاً تمامی نکات و فرض‌هایی که برای پیاده‌سازی‌ها و محاسبات خود در نظر می‌گیرید را در گزارش ذکر کنید و تمامی اصول نگارشی را مطابق با فایل ارسالی در صفحه درس رعایت بفرمایید.
- الزامی به ارائه توضیح جزئیات کد در گزارش نیست. اما باید نتایج بدست آمده را گزارش و تحلیل کنید.
- برای پیاده‌سازی می‌توانید از محیط Colab استفاده نمایید.
- لطفا گزارش (درقالب PDF) ، فایل کدها و سایر ضمائم مورد نیاز را با فرمت زیر در صفحه درس در سامانه یادگیری الکترونیکی بارگذاری نمائید.

HW#_[Lastname]_[StudentNumber].zip

- در صورت وجود هرگونه ابهام یا مشکل می‌توانید بر اساس شماره سوال از طریق رایانامه‌های زیر با دستیاران آموزشی مربوطه در تماس باشید.
  - **سوال اول:**
    - حسین پورمهرانی : h.pourmehrani@gmail.com
  - **سوال دوم:**
    - پیمان باقرشاهی : p.baghershahi@ut.ac.ir