

به نام خدا



دانشگاه تهران



دانشکده مهندسی برق و کامپیوتر

درس شبکه‌های عصبی و یادگیری عمیق

تمرین ششم

نام دستیار طراح	محمدامین غنی زاده	پرسش ۱
رایانامه	ghanizadeh.amin@ut.ac.ir	
نام دستیار طراح	محمد گرجی	پرسش ۲
رایانامه	mohamadgorjicode@gmail.com	
مهلت ارسال پاسخ	۱۴۰۳.۰۴.۰۳	

فهرست

۱	قوانین
۱	پرسش ۱. Variational Auto-Encoder
۱	معرفی دیتاست
۱-۱	پیش پردازش دیتاست
۱-۲	ساخت VAE روی دیتاست ها
۳-۱	استفاده از یک مدل برای دو دیتاست
۴-۱	VQ-VAE
۴-۱-۲	VQ-VAE 2.۵
۵	پرسش ۲. Image Translation
۵-۱-۲	آشنایی با Image Translation و معماری Pix2Pix
۵-۲-۲	پایه سازی معماری Pix2Pix

شکل‌ها

- شکل ۱. نمونه‌هایی از دیتاست Anime Faces ۱
- شکل ۲. نمونه تولیدی مدل از Anime Faces ۲
- شکل ۳. نمونه تولیدی مدل از Cartoon Faces ۲
- شکل ۴. یک نمونه از عکس ماهواره‌ای و تصویر نقشه متناظر آن 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... انجام دهید)
- **کیفیت گزارش شما در فرآیند تصحیح از اهمیت ویژه‌ای برخوردار است؛** بنابراین، لطفا 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 حتما باید در قالب نوت‌بوک با پسوند ipynb تهیه شوند، در پایان کار، تمامی کد اجرا شود و خروجی هر سلول حتما 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. در صورتی که دو گروه از یک منبع مشترک استفاده کنند و کدهای مشابه تحویل دهند، تقلب محسوب می‌شود.
- نحوه محاسبه تاخیر به این شکل است: پس از پایان رسیدن مهلت ارسال گزارش، حداکثر تا یک هفته امکان ارسال با تاخیر وجود دارد، پس از این یک هفته نمره آن تکلیف برای شما صفر خواهد شد.

○ سه روز اول: بدون جریمه

○ روز چهارم: ۵ درصد

○ روز پنجم: ۱۰ درصد

○ روز ششم: ۱۵ درصد

○ روز هفتم: ۲۰ درصد

- حداکثر نمره‌ای که برای هر سوال می‌توان اخذ کرد ۱۰۰ بوده و اگر مجموع بارم یک سوال بیشتر از ۱۰۰ باشد، در صورت اخذ نمره بیشتر از ۱۰۰، اعمال نخواهد شد.

○ برای مثال: اگر نمره اخذ شده از سوال ۱ برابر ۱۰۵ و نمره سوال ۲ برابر ۹۵ باشد، نمره نهایی تمرین ۹۷.۵ خواهد بود و نه ۱۰۰.

- 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.

پیش ۱. Variational Auto-Encoder

هدف این سوال آشنایی با متد variational autoencoder و تولید تصویر با استفاده از این متد است. سپس در بخش دوم، conditional variational autoencoder ها و تفاوت آنها با VAE های معمولی را بررسی می‌کنیم. در نهایت متد VQ VAE را بررسی کرده و پیاده سازی خواهیم کرد.

معرفی دیتاست

دیتاست های مورد استفاده برای این سوال، دیتاست [Anime Face Dataset](#) و دیتاست [Cartoon Faces](#) خواهند بود. این دیتاست ها را می‌توانید از Kaggle دانلود کنید. پیشنهاد می‌شود برای راحتی کار، کد خود را داخل colab یا Kaggle نوشته و اجرا کنید.



شکل ۱. نمونه‌هایی از دیتاست Anime Faces

۱-۱. پیش پردازش دیتاست

دیتاست های معرفی شده در بالا را دانلود کرده و پیش پردازش های لازم را روی آنها انجام دهید. توجه داشته باشید که از ۵ تا از فولدرهای داخل Cartoon faces را به صورت رندوم انتخاب کرده و استفاده کنید. همچنین برای راحتی کار برای بخش CVAE می‌توانید عکس‌های دو دیتاست را داخل یک فولدر جدا کپی کنید، فقط توجه داشته باشید که لیبل های دو دیتاست برای آموزش CVAE در بخش دوم باید متفاوت باشند.

۱-۲. ساخت VAE روی دیتاست ها

(۵۰ نمره)

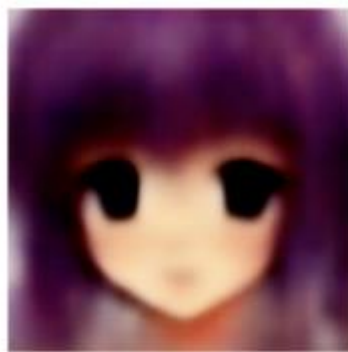
ابتدا توضیحاتی کلی از اینکه مدل های VAE چگونه کار می‌کنند ارائه دهید و لاس به کار رفته در آن را توضیح دهید. (۳ نمره)

دیتاست Anime Face را لود کرده و ۵ عکس تصادفی از آن را نمایش دهید. (۲ نمره)

سپس یک مدل با معماری کانولوشن را ساخته و مدل را روی این دیتاست آموزش دهید. آموزش را حداقل سه اپایک ادامه دهید. (جزئیات معماری مانند هایپرپارامترها و تعداد لایه‌ها بر عهده خودتان است. همچنین می‌توانید تصاویر را برای افزایش سرعت آموزش، به مقدار دلخواه resize کنید اما در این بخش از همه چنل‌ها استفاده کنید. همچنین تصاویر را به اندازه‌های خیلی کوچک resize نکنید. دقت کنید که مدل تان از لایه‌های کانولوشن تشکیل شود) (۱۰ نمره)

پس از اتمام آموزش، Reconstruction Loss و لاس کلی را طی آموزش نمایش دهید. سپس ۸ عدد نویز تولید کرده و با استفاده از این نویزها، تصویر تولید کرده و آنها را نمایش دهید. (۱۰ نمره)

نمونه‌ای از عکسی که ممکن است تولید شود را در زیر مشاهده می‌کنید.



شکل ۲. نمونه تولیدی مدل از Anime Faces

سپس، می‌خواهیم مدلی را روی Cartoon faces آموزش دهیم. این دیتاست شامل چندین فولدر است، همانطور که در بخش پیش پردازش اشاره شد، ۵ فولدر را به صورت رندوم انتخاب کرده و به عنوان دیتاست از عکس‌های داخل آنها استفاده کنید. سپس همه‌ی مراحل که در بالا انجام دادید را با همان مدل (دقت کنید که نه همان وزن‌های بخش قبلی، وزن‌ها را مجدداً initialize کنید) روی دیتاست Cartoon Faces تکرار کنید. نمونه خروجی مدل برای این دیتاست در زیر آمده است:



شکل ۳. نمونه تولیدی مدل از Cartoon Faces

۳-۱. استفاده از یک مدل برای دو دیتاست

(۳۰ نمره)

ابتدا، در مورد Conditional VAE ها تحقیق کرده و توضیح مختصری در مورد آنها و تفاوت آنها با VAE ساده بدهید. (۵ نمره)

در بخش قبلی، همان طور که دیدید برای هر کدام از دیتاست ها، مدلی جدا آموزش دادیم، در این بخش می‌خواهیم یک مدل را روی هر دو دیتاست آموزش دهیم و کار تولید را با استفاده از یک مدل انجام دهیم. برای این بخش، از نصف داده‌هایی که در بخش قبلی استفاده کردید استفاده کنید. به این معنی که نصف دیتاست Anime faces و نصف ۵ عدد فولدری که برای Cartoon faces استفاده کردید، دیتاست این بخش را تشکیل می‌دهد.

سپس یک مدل Conditional VAE بسازید و روی همه‌ی این دیتاست آموزش دهید. دقت کنید که کلاس‌های هر کدام از دیتاست ها باید لیبل جدا داشته باشند. (به عنوان مثال، به عکس‌های Anime Faces لیبل ۰ و به Cartoon faces لیبل ۱ بدهید). همچنین برای این قسمت، می‌توانید فقط از یک چنل تصاویر برای آموزش استفاده کنید و همچنین در استفاده از هر کدام از معماری‌ها آزاد هستید. (می‌توانید

از کانولوشن استفاده نکنید) (۱۰ نمره)

پس از اتمام آموزش، یک بار با دادن لیبل‌هایی که به دیتاست Anime Faces دادید، ۸ تصویر از این کلاس تولید کنید و بار دیگر با دادن لیبل مختص Cartoon faces، ۸ تصویر از آن تولید کنید. (تصاویر تولیدی باید توسط یک مدل باشند) (۱۵ نمره)

توجه کنید که برای این قسمت، کیفیت تصاویر تولیدی خیلی مهم نیست ولی مدل باید بتواند از روی لیبل داده شده، تصویر مرتبط با آن لیبل را تولید کند.

۴-۱. VQ-VAE

(۲۰ نمره)

[این مقاله](#) را بخوانید و توضیح مفصلی از عملکرد آن و تفاوت‌های آن با VAE ساده بدهید. (۵ نمره)

سپس این مدل را پیاده سازی کرده و روی دیتاست Anime Faces آموزش دهید. پس از اتمام آموزش، Reconstruction loss در حین آموزش را رسم کنید. (۱۰ نمره)

سپس، ۸ تا از عکس‌های Anime Faces را انتخاب کرده و به مدل بدهید و خروجی را نمایش دهید. دقت کنید که در این بخش لازم نیست از روی نویز، عکس تولید کنید و فقط تصاویر موجود را به مدل می‌دهید. همچنین ۸ تصویری را هم که در بخش اول از دیتاست Anime Faces تولید کردید را به مدل داده و خروجی را نمایش دهید. (۵ نمره)

۲.۵-۱. VQ-VAE

(۵ نمره امتیازی)

مقاله [VQ-VAE-2](#) را مطالعه کرده و توضیح دهید.

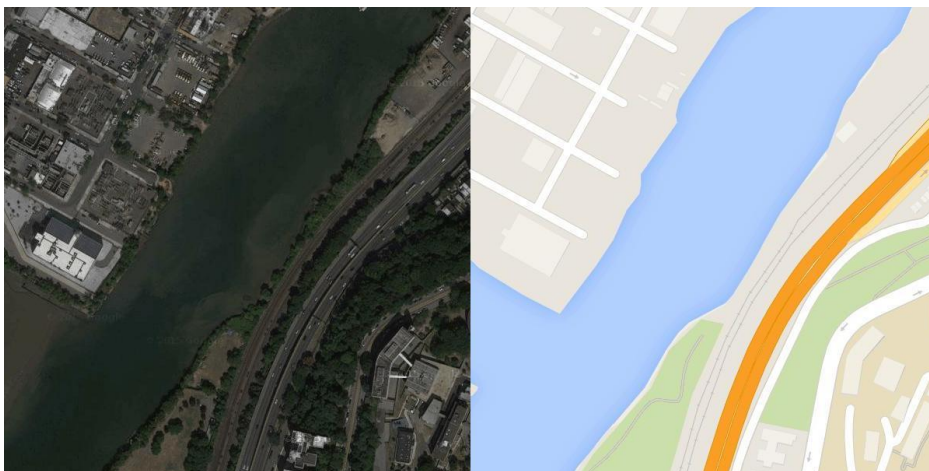
پ.ن: دقت شود که برای این تمرین، بخش زیاد نمره به تصاویری که تولید می‌کنید داده می‌شود، پس در تولید تصاویر خواسته شده دقت فرمایید.

۱-۲. آشنایی با Image Translation و معماری Pix2Pix

Image Translation یک حوزه از بینایی ماشین است که هدف آن به دست آوردن یک نگاشت از تصویر اولیه به تصویر ثانویه است. از مصادیق کاربردی این حوزه می‌توان به موارد زیر اشاره نمود:

- ۱- انتقال سبک: تغییر سبک تصویر با حفظ محتوا (تبدیل عکس به نقاشی با یک سبک خاص)
- ۲- داده افزایی: تولید تصاویر جدید بر اساس تصاویر آموزشی موجود برای بهتر پوشش دادن توزیع دادگان هنگام آموزش (معمولا با تغییر در یک یا چند ویژگی برای حفظ ناشناس بودن و یا کاهش یک سوگیری خاص در مجموعه دادگان نیز استفاده می‌گردد)
- ۳- بازیابی تصویر: کاهش نویز، افزایش وضوح، رنگ آمیزی و بازیابی قسمت‌های مخدوش تصاویر
- ۴- تطبیق دامنه: تبدیل تصاویر از یک دامنه خاص به دامنه‌ای دیگر (تبدیل خطوط طراحی به عکس)

در حوزه شبکه‌های مولد متخاصم از دو مدل CycleGAN و Pix2Pix می‌توان به عنوان معماری‌های مطرح برای انجام چنین کارکردی اشاره کرد. در این بخش از تمرین می‌خواهیم با استفاده از معماری Pix2Pix عکس‌های ماهواره‌ای از سطح شهر را به تصاویری تبدیل کنیم که گویای نقشه متناظر با این تصویر باشند. یک نمونه از چنین تناظری در شکل ۴ قابل مشاهده است.



شکل ۴. یک نمونه از عکس ماهواره‌ای و تصویر نقشه متناظر آن

بدین منظور ابتدا مقاله [Pix2Pix](#) را مطالعه کرده و به سوالات زیر پاسخ دهید:

۱- ابتدا بیان کنید که تفاوت‌های اصلی این مدل با یک مدل GAN ساده در چیست؟ (از منظر مفهومی و نه جزئیات معماری) و سپس یک کاربرد دیگر برای چنین مدلی برای انجام یک تسک Image Translation در حوزه تصاویر پزشکی بیان کنید (پاسخ می‌تواند به ایده خودتان و یا با ارجاع به یک مقاله مشخص باشد). (۵ نمره)

۲- تفاوت‌های بخش Discriminator این مدل را با Discriminator یک GAN عادی به همراه شکل بیان کنید. همچنین مطرح کنید که این تفاوت چگونه می‌تواند در راستای بهبود کلی عملکرد مدل عمل کند. (۱۰ نمره)

۳- بخش Generator این مدل از معماری U-Net استفاده می‌کند، همراه با شکل جزئیات این بخش را شرح دهید (نحوه کارکرد، اجزا مختلف و ابعاد ورودی و خروجی هر بلاک). همچنین بیان بدارید که کارکرد DropOut در تنوع بخشی به نتایج چیست؟ (۱۵ نمره)

۴- تابع هزینه نهایی این معماری را بر اساس بیانات مقاله تفسیر کنید و توضیح دهید که با توجه به تابع هزینه، بروزرسانی وزن‌های هر زیربخش از معماری کلی به چه صورت انجام می‌پذیرد. (۱۰ نمره)

۵- دو نمونه از پژوهش‌های مرتبط با معماری Pix2Pix در راستای بهبود کیفیت مدل را به همراه نام مقاله ذکر کنید و ایده هر کدام را در یک الی دو پاراگراف توضیح دهید. (به عنوان مثال Pix2PixHD) (۱۰ نمره)

۲-۲. پیاده سازی معماری Pix2Pix

۱- ابتدا فولدر دادگان maps را از [لینک مورد نظر](#) دریافت کنید و یک مجموعه داده متشکل از دادگان آموزش و اعتبارسنجی آن بسازید. سپس ۳ نمونه از تصاویر را به انتخاب خودتان نمایش دهید. (۱۰ نمره)

۲- با توجه به سوال ۲ بخش قبلی کلاس Discriminator خود را نیز تعریف کنید. همچنین با توجه به سوال ۳ بخش قبلی کلاس‌های DownSample و UpSample و در انتها کلاس Generator که خود متشکل از دو جزء قبلی است را پیاده سازی کنید. (۱۵ نمره)

۳- با توجه به سوال ۴ بخش قبل توابع هزینه مدل را پیاده سازی کنید و مدل را در ۲۰ دوره (epoch) آموزش دهید (۱۰ دوره نیز قابل قبول است). همچنین نمودار توابع هزینه Generator و Discriminator را رسم و تحلیل نمایید و مقدار ابرپارامترهای مورد استفاده را ذکر کنید. (۱۵ نمره)

- ۴- پس از هر دوره آموزشی تعدادی از تصاویر تولید شده را (به عنوان مثال ۵ عکس) به همراه دو عکس متناظر آن (عکس ماهواره‌ای و تصویر نقشه) نمایش دهید و نتیجه ایپاک آخر را در گزارش خود بیاورید. (۱۰ نمره)
- ۵- مدل را به روی یک مجموعه داده دیگر از لینک قرار داده شده و یا یک مجموعه دلخواه دیگر آموزش دهید و نتایج را به مانند بخش‌های پیشین تحلیل کنید. (۵ نمره امتیازی)