



پروژه ششم

هدف: آشنایی با شبکه های مولد تقابلی و ترجمه تصویر به تصویر.

کد: پیاده سازی این پروژه را به زبان پایتون انجام دهید؛ در این فعالیت مجاز به استفاده از tensorflow یا pytorch می باشید.

گزارش: ملاک اصلی انجام فعالیت، گزارش آن است و ارسال کد بدون گزارش فاقد ارزش است. برای این فعالیت یک فایل گزارش در قالب pdf تهیه کنید و در آن برای هر سوال، تصاویر ورودی، تصاویر خروجی و توضیحات کامل و جامعی تهیه کنید. **تذکر:** مطابق قوانین دانشگاه هر نوع کپی برداری و اشتراک کار دانشجویان غیر مجاز بوده و شدیداً برخورد خواهد شد. استفاده از کدها و توضیحات اینترنت به منظور یادگیری بلامانع است، اما کپی کردن غیرمجاز است.

راهنمایی: در صورت نیاز میتوانید سوالات خود را در خصوص پروژه از تدریسارهای درس، از طریق ایمیل زیر یا در گروه تلگرامی بپرسید. (لینک گروه تلگرامی در سایت کورسز در دسترس بوده و قبلاً به همه ی دانشجویان ایمیل شده است)

Email: ann.ceit.aut@gmail.com

توجه: برای آموزش شبکه های عمیق می توانید از منابع و بسترهای سخت افزاری برخط رایگان نظیر Google Colab یا Kaggle استفاده نمایید.

تاخیر مجاز: در طول ترم، مجموعاً مجاز به حداکثر ده روز تاخیر برای ارسال تمرینات هستید (بدون کسر نمره). این تاخیر را می توانید بر حسب نیاز بین تمرینات مختلف تقسیم کنید؛ اما مجموع تاخیرات تمام تمرینات شما نباید بیشتر از ده روز شود. پس از استفاده از این تاخیر مجاز، هر روز تاخیر باعث کسر ۱۰٪ نمره ی آن تمرین خواهد شد.

ارسال: فایل های کد و گزارش خود را در قالب یک فایل فشرده با فرمت StudentID_HW06.zip تا تاریخ ۱۴۰۲/۰۴/۰۷ ارسال نمایید.

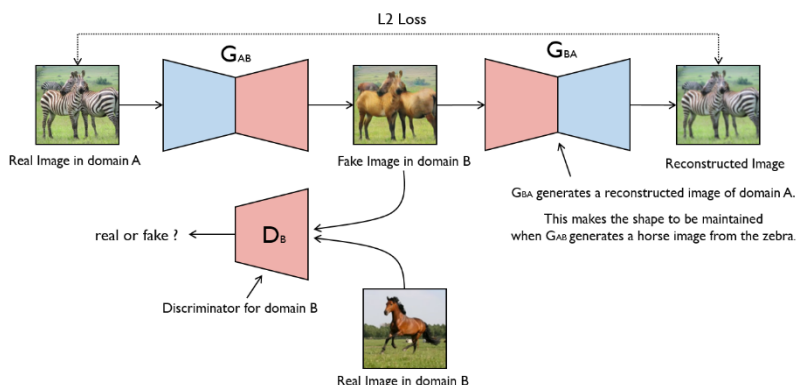
ترجمه تصویر به تصویر^۱ به مبحثی از حوزه بینایی کامپیوتر اطلاق می شود که ورودی و خروجی مدل هر دو تصویر می باشد. گاه تصویر خروجی کلاً یک تصویر دیگر می باشد مانند تولید نقشه ارتباطی شهری بر اساس تصویر هوایی؛ همچنین گاه تصویر خروجی بر پایه ی تصویر ورودی بوده اما برخی اصلاحات یا تغییرات معنا ایجاد می شود مانند تغییر رنگ پوست انسان یا افزودن/حذف عینک از چهره. در این پروژه هدف استفاده از شبکه های مولد تقابلی^۲ با محوریت ترجمه تصویر به تصویر می باشد.

^۱ Image-To-Image Translation

^۲ Generative Adversarial Networks (GANs)

شما در کلاس درس با ایده‌ی شبکه‌های مولد تقابلی و تابع هزینه آن بصورت کامل آشنا شده اید. یک معماری که با عنوان CycleGAN^۳ شناخته می‌شود، برای ترجمه‌ی تصویر به تصویر می‌تواند مورد استفاده قرار گیرد. طرحواره‌ی این معماری در شکل (۱) آورده شده است؛ این شبکه از دو شبکه‌ی خودکدگذار تغییراتی^۴ به عنوان مولد^۵ و یک شبکه‌ی عمیق پیچشی^۶ به عنوان تمایزگر برای هدف مذکور استفاده می‌کند.

۱. ساختار و چگونگی عملکرد شبکه‌ی CycleGAN را ضمن فرآیند آموزش و آزمون توضیح داده و بیان کنید که اهداف هر کدام از شبکه‌های خودکدگذاری تغییراتی چیست و به چه نحوی در امر ترجمه‌ی تصویر به تصویر کمک می‌کند. همچنین به طور خلاصه تابع هزینه‌های مورد استفاده و استدلال استفاده از آنان را شرح دهید. (۲۰ نمره)



۲. در آموزش شبکه‌های مولد تقابلی، چندین مشکل مرسوم وجود دارد. سه مورد از آنان بترتیب حالت سقوط^۷، کمبود تنوع در داده‌های تولیدی^۸ و متعادل‌سازی آموزش شبکه‌های مولد و تمایزگر است. سه مشکل مذکور را بطور خلاصه مورد بررسی قرار داده و برای هر یک راهکارهای پیشنهاد شده در سطح صنعت/آکادمیک را بیان کنید. (۱۲ نمره)

۳. آیا آموزش شبکه‌ی CycleGAN با نظارت است یا بدون نظارت؟ (۳ نمره)

برای این پروژه مجموعه داده‌ی summer2winter از مجموعه داده‌های دانشگاه Berkeley در اختیار شما قرار گرفته است که لینک دسترسی در انتهای متن تمرین قرار گرفته است. مجموعه داده‌ی مذکور بصورت پیشفرض به دو دسته‌ی آزمون و آموزش تقسیم شده و نیاز به تفکیک از جانب شما نمی‌باشد. هدف این مجموعه داده ترجمه تصویر به تصویر با محوریت تغییر فصل در تصویر ورودی از تابستان/بهار به فصل زمستان می‌باشد بطوری که عکسی از فصل تابستان/بهار را ورودی بگیرد که در آن منظره و صحنه سرسبز است و عکس متناظری با همان صحنه و منظره خروجی دهد با این تفاوت که سرسبزی از بین رفته و برف باریده و احساس زمستان در تصویر حاکم است.

۴. شبکه‌ی CycleGAN را با هدف ترجمه‌ی تصویر به تصویر برای مجموعه داده‌ی مد نظر پیاده سازی کرده و فرآیند آموزش را به همراه معماری نهایی هر کدام از سه شبکه گزارش نمایید. همچنین نمودار خطای کل شبکه و نمودار خطای هر کدام از زیر شبکه‌ها را بصورت مجزا رسم نمایید. (۵۰ نمره)

۵. از مجموعه داده‌ی آزمون ۲۰ تصویر بصورت تصادفی انتخاب کرده و به همراه ترجمه‌ی تصویری‌شان در گزارش بیاورید. همچنین از فضای شهری و دانشگاه که سرسبزی در آن نمایان است، ۵ عکس مختلف گرفته و برای هر یک فرآیند مذکور را تکرار

³ Zhu, J.Y., Park, T., Isola, P. and Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2223-2232).

⁴ Variational Auto Encoders (VAE)

⁵ Generator

⁶ Convolutional Neural Networks (CNN)

⁷ Mode Collapse

⁸ Lack of diversity

کنید. در برخی عکس‌ها ممکن است نتایج رضایت بخش نباشد و دارای برخی مشکلات جزئی و کلی باشند، ضمن بررسی این مشکلات بنظر شما دلیل چیست و چگونه می‌توان آن را حل نمود؟ (۱۵ نمره)

۶. برای بهبود نتایج و آموزش مدلی بهتر می‌توانید از روش‌های افزونگی تصویر^۹ استفاده نموده و تعداد نمونه‌های آموزش را افزایش دهید. ضمن توضیح این رویکرد و بررسی مختصر روش‌های این حوزه، آن را پیاده‌سازی کرده و در یک شبکه‌ی یکسان نتیجه را مقایسه کنید. لازم به ذکر است برای مشاهده‌ی فرق در نتایج باید شبکه‌ها تا حد امکان صحیح طراحی شده باشند و تکرار آموزش به اندازه کافی بوده باشد. (۱۵ امتیاز تشویقی)

توجه: برای کاهش پیچیدگی محاسبات و سنگینی شبکه‌ها می‌توانید ابعاد تصاویر را حداکثر تا 64×64 کاهش دهید.

لینک مجموعه داده:

https://people.eecs.berkeley.edu/~taesung_park/CycleGAN/datasets/summer2winter_yosemite.zip

موفق باشید

⁹ Image Augmentation