

به نام خدا



دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده برق و کامپیوتر



شبکه‌های عصبی و یادگیری عمیق

مینی پروژه شماره ۳

تیر ۹۹

فهرست سؤالات

۳.....	توضیحات کلی تمرین
۵.....	سؤال ۱ – Variational Autoencoder
۷.....	سؤال ۲ – DCGAN
۸.....	سؤال ۳ – Conditional GANs
۹.....	سؤال ۴ – کاربردهای GAN (امتیازی)

توضیحات کلی مینی پروژه

در این تمرین هدف آشنایی شما با Generative adversarial Network و Variational Autoencoder است. در سؤال یک از شما پیاده‌سازی VAE و کار با آن خواسته می‌شود، در سؤال دوم شما با یکی از متداول‌ترین معماری‌های GAN یعنی DCGAN آشنا می‌شوید و باید آن را پیاده‌سازی کنید، در سؤال سوم با معماری‌های Conditional GAN آشنا خواهید شد که به بررسی ۴ معماری معروف از آن‌ها پرداخته می‌شود که شما از بین آن‌ها باید CGAN را پیاده‌سازی کنید و از بین معماری‌های InfoGAN، Semi-Supervised-GAN و ACGAN نیز به انتخاب خودتان یکی از آن‌ها را مطالعه و پیاده‌سازی کنید.

در سؤال آخر نیز با کاربردهای جذاب GAN ها آشنا می‌شوید و از شما اجرای یکی از آن‌ها خواسته می‌شود، این سؤال امتیازی و نمره مثبت است.

با توجه به اینکه آموزش GAN ها به‌طور کلی نیاز به سخت‌افزار قدرتمندی دارد، پیشنهاد می‌شود از محیط Google Colab و تحت GPU استفاده کنید یا از یک GPU قدرتمند دیگر استفاده کنید. در هنگام استفاده از محیط Google Colab برای اینکه آموزش شبکه‌ها برای شما سریع‌تر انجام شود، بررسی کنید که GPU ای که Google Colab در اختیار شما قرار می‌دهد از نوع P100 باشد.

از آنجایی که سیستمی که در اختیار دانشجویها است ممکن است متفاوت باشد، از شما در این تمرین اجرای تعداد مشخصی اپیاک خواسته نمی‌شود تا دچار مشکل سخت‌افزاری نشوید، پس با توجه به سیستمی که در اختیار دارید تعداد اپیاک های آموزشی را خودتان تنظیم کنید، در هیچ‌کدام از سؤال‌ها انتظار نمی‌رود بیش از ۲ ساعت برای آموزش شبکه تحت GPU وقت بگذارید و هدف به هیچ‌عنوان رسیدن به دقتی برابر با دقت مقاله‌ها نیست. مهم این است که پیاده‌سازی شبکه به‌درستی انجام شده باشد و با افزایش اپیاک های آموزشی نتیجه و خروجی شبکه بهبود یابد. نمونه‌ای از تصاویری که انتظار می‌رود با سخت‌افزاری مشابه Google Colab به دست آورید در قسمت پیوست قرار داده شده است.

در تمام سؤال‌ها از شما توضیح سازوکار معماری و همچنین نتایج پیاده‌سازی آن خواسته می‌شود که با توجه به مشترک بودن آن‌ها در ادامه توضیح داده می‌شود که چطور این سؤالات را جواب بدهید.

توضیح سازوکار معماری:

برای هر معماری که در تمرین از شما پیاده‌سازی آن خواسته شده است، لازم است مقاله مربوط به آن معماری را مطالعه کنید و در حدود چند پاراگراف یا یک صفحه در مورد **سازوکار معماری** آن توضیح دهید.

نتایج موردنیاز بعد از اجرای هر مدل:

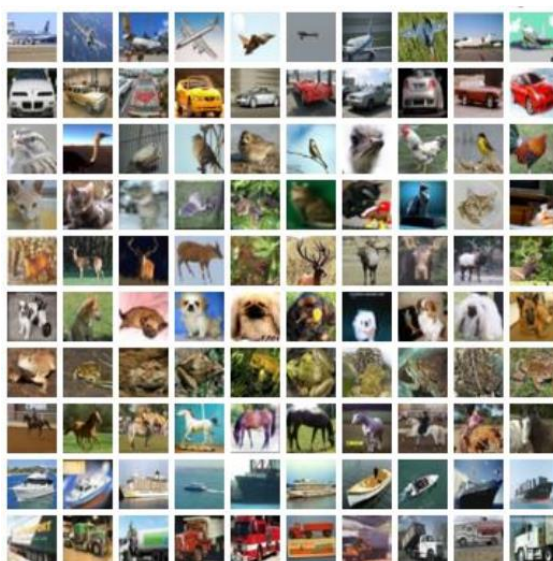
۱ - نمونه تصاویر تولیدشده:

برای هر مدلی که در سؤالات برای تولید داده‌ها استفاده می‌کنید نتایج را به این شکل برای ۵ ایپاک متفاوت نمایش دهید، انتخاب اینکه در چه ایپاک هایی این تصاویر را نمایش دهید بستگی به مدت‌زمان آموزش مدل شما و تعداد ایپاک هایی اجرایی شما دارد و بر عهده خود شما است، سعی کنید از ایپاک های ابتدایی تا انتهای تعدادی ایپاک را برای نمایش تصاویر انتخاب کنید که روند پیشرفت آموزش شبکه و بهبود کیفیت تصاویر تولیدی را به‌خوبی نشان دهد.

تصاویر تولیدی باید به‌صورت یک ماتریس ۱۰ در ۱۰ به‌صورت زیر باشند، برای این منظور می‌توانید از کتابخانه Pillow یا سایر کتابخانه‌ها استفاده کنید.



تصویر ۲ (نمونه تصاویر تولیدی برای مجموعه داده MNIST)



تصویر ۱ (نمونه تصاویر تولیدی برای مجموعه داده CIFAR10)

۲ - نمودار دقت یا Loss:

بسته به مدل شبکه انتخابی در تمرین‌ها و نوع Loss مورد استفاده، نمودار دقت یا Loss را به ازای هر ایپاک رسم کنید.

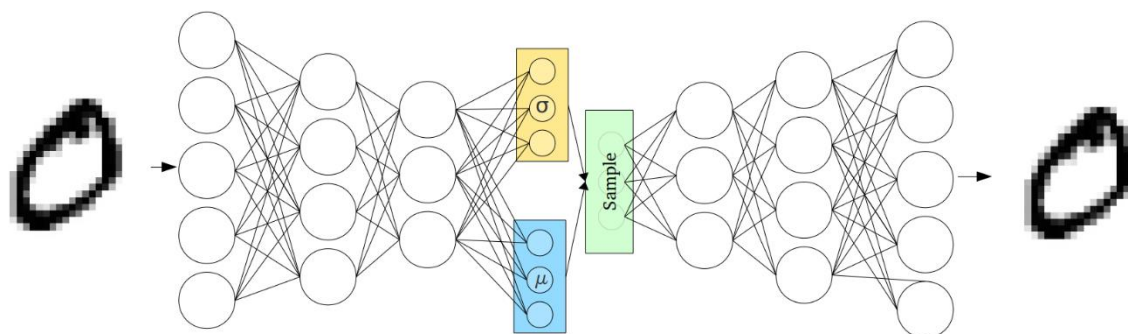
توضیح کد:

همچنین از شما در سؤالات خواسته می‌شود که روند پیاده‌سازی خود را در فایل Notebook کد توضیح دهید. توصیه می‌شود برای هر مدل کد خود را در یک فایل Notebook بنویسید و روند پیاده‌سازی را همراه با کدها توضیح دهید، به هیچ‌عنوان نیازی به ذکر جزئیات نیست و روند کلی کفایت می‌کند. برای توضیح کد خود در فایل Notebook یک بخش text ایجاد کنید و توضیحات خود را برای هر قسمت یادداشت نمایید. این توضیح می‌تواند به زبان فارسی یا انگلیسی باشد.

با توجه به این توضیحات دیگر نیازی به ذکر روند و چالش‌های پیاده‌سازی مدل در گزارش نیست.

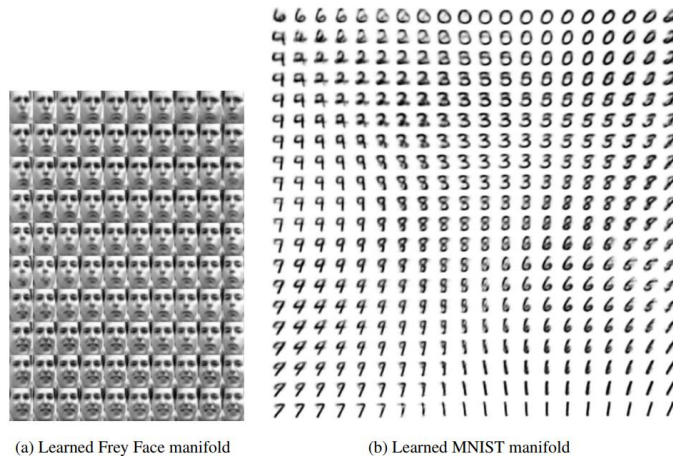
سؤال ۱ – Variational Autoencoder

شبکه‌های Variational Autoencoder که به اختصار VAE نامیده می‌شود، جز مدل‌های Generative قدرتمندی هستند که در تولید داده‌ها از آن‌ها استفاده می‌شود، این شبکه‌ها مانند سایر Autoencoder ها از دو بخش Encoder و Decoder تشکیل شده است.



تصویر ۳) یک شمای کلی از VAE

تفاوت اصلی شبکه‌های Variational Autoencoder با Autoencoder های معمولی در این است که در بخش Encoder برخلاف روش قبلی، دو بردار تولید می‌شود، یک بردار شامل تعدادی میانگین و یک بردار شامل تعدادی انحراف معیار است؛ و در بخش Decoder پس از نمونه‌برداری از بردارهای میانگین و انحراف معیار خروجی تولید می‌شود. از نظر خروجی نیز شبکه‌های Autoencoder سعی می‌کنند دقیقاً تصویری مانند تصویر ورودی تولید کنند در حالی که خروجی‌های تصاویر Variational Autoencoder دقیقه همان تصویر نیستند و تصاویر جدیدی می‌توانند باشند. مقاله VAE را از [اینجا](#) می‌توانید دانلود کنید. نمونه‌ای از خروجی VAE را برای دو مجموعه داده در تصاویر زیر مشاهده نمایید.



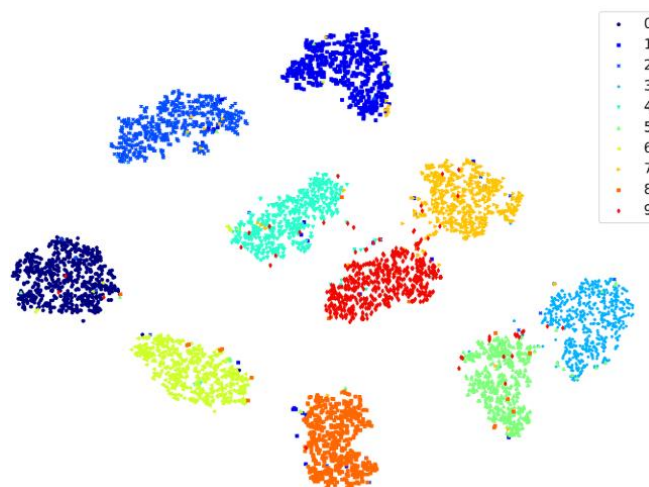
تصویر ۴) نمونه‌ای از تصاویر تولیدشده توسط VAE

الف) تفاوت Generative Adversarial network و Variational Autoencoder را در تولید داده‌ها بیان کنید و مزایا و معایب آن‌ها را نسبت به یکدیگر بیان نمایید.

ب) مدل VAE را با استفاده از دیتاست MNIST پیاده‌سازی کنید و نتایج را مانند آنچه در صفحه ۴ خواسته‌شده در گزارش بیاورید. مراحل پیاده‌سازی VAE را داخل فایل notebook کد توضیح دهید. برای ساختار لایه‌های Encoder و Decoder محدودیت خاصی وجود ندارد و می‌توانید از معماری موردنظر خودتان استفاده کنید ولی سعی کنید تا حد امکان به مطالب نوشته‌شده در مقاله VAE وفادار باشید.

ج) در قسمت ب از چه تابع هزینه‌ای استفاده کردید؟ چرا؟

د) به کمک Encoder، تصاویر داده‌های Validation را از فضای ورودی ۷۸۴ بعدی به فضای دوبعدی برده و آن‌ها را Cluster بندی کنید و نتایج را مانند تصویر زیر گزارش کنید.



تصویر ۵) خوشه‌بندی کردن داده‌های اعداد صفر تا ۹ در فضای دوبعدی

(امتیازی)

ه) معماری EfficientNet یکی از **جدیدترین** معماری‌های کانولوشنی ارائه شده است که نسبت به سایر مدل‌های ارائه شده تاکنون، توانسته است در بیشتر دیتاست‌های تصویری **دقت بالاتری** را کسب کند، این در حالی است که تعداد پارامترهای مورد استفاده این معماری از سایر معماری‌ها نیز کمتر است که این مسئله منجر به افزایش سرعت آموزش این معماری و کاهش نیاز سخت‌افزاری شده است.

حال با استفاده از Transfer Learning و مبتنی بر مدل EfficientNet قسمت (ب) این سؤال را مجدداً پیاده‌سازی کنید.

سؤال ۲ – DCGAN

در این سؤال و سؤالات بعدی از شما خواسته می‌شود که معماری‌های مختلف شبکه عصبی مصنوعی بر پایه GAN را برای تولید داده‌ها پیاده‌سازی نمایید. با توجه به اینکه GAN ها بر پایه دو شبکه مفهومی مجزا به اسم **Generator** و **Discriminator** به صورت adversarial کار می‌کنند. روش‌های متفاوتی تاکنون برای طراحی معماری داخل این دو شبکه معرفی شده است که هر کدام از آن‌ها می‌توانند کاربردهای متفاوتی داشته باشند.

یکی از معروف‌ترین این مدل‌ها مدل DCGAN است که در این سؤال از شما پیاده‌سازی آن خواسته شده است، مقاله مدل DCGAN را از [اینجا](#) می‌توانید مشاهده نمایید.



تصویر ۶) نمونه تصاویر تولیدشده توسط GAN

الف) سازوکار طراحی معماری DCGAN را شرح دهید.

ب) نحوه ایجاد **نویز** را توضیح دهید.

(ج) دلیل استفاده از Strided Convolution و Fractional Strides Convolution را توضیح دهید.

(د) مدل DCGAN را با استفاده از دیتاست CIFAR10 پیاده‌سازی کنید و نتایج را مانند آنچه در صفحه ۴ خواسته شده در گزارش بیاورید. مراحل پیاده‌سازی DCGAN را داخل فایل notebook کد توضیح دهید.

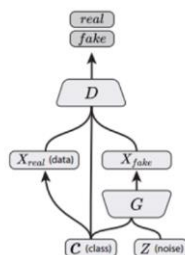
(ه) یکی از چالش‌ها در شبکه‌های GAN، چالش انتخاب تابع Loss مناسب است (با توجه به اینکه این شبکه‌ها از دو قسمت Generator و Discriminator تشکیل شده‌اند)، در مورد این مسئله تحقیق کنید و ۴ تابع هزینه معروف برای این کار را نام ببرید و مختصراً توضیح دهید.

راهنمایی: برای قسمت (ه) در مورد Wasserstein GAN و Least Squares GAN تحقیق کنید.

سؤال ۳ – Conditional GANs

در این بخش به بررسی Conditional GAN ها می‌پردازیم، این مدل معماری‌ها برخلاف معماری‌های سؤال قبل تنها از خود تصاویر استفاده نمی‌کنند، بلکه از یکسری **اطلاعات اضافی** نیز برای تولید تصاویر استفاده می‌کنند.

در سال ۲۰۱۴ ارائه شده است، از Label هر تصویر به عنوان اطلاعات اضافه در شبکه GAN استفاده شده است تا بتوان کیفیت تصاویر تولیدی را افزایش داد. مدل معماری این شبکه را در تصویر زیر مشاهده می‌نمایید.



تصویر ۷) معماری کلی CGAN

مقاله CGAN را از [اینجا](#) می‌توانید دانلود کنید.

(الف) سازوکار طراحی معماری CGAN را شرح دهید.

(ب) مدل CGAN را با استفاده از دیتاست CIFAR10 پیاده‌سازی کنید و نتایج را مانند آنچه در صفحه ۴ خواسته شده در گزارش بیاورید. مراحل پیاده‌سازی DCGAN را داخل فایل notebook کد توضیح دهید.

(ج) از جمله دیگر شبکه‌هایی که از این دسته به شمار می‌روند می‌توان به [InfoGAN](#)، [AC-GAN](#) و [Semi-Supervised GAN](#) اشاره کرد، یکی از آن‌ها را **به دلخواه** انتخاب کرده، سازوکار معماری آن را توضیح داده و مانند قسمت ب آن را پیاده‌سازی نمایید.

سؤال ۴ – کاربردهای GAN (امتیازی)

معرفی:

در این بخش قسمتی دیگر از کاربردهای شبکه‌های GAN رو بیان می‌کنیم.

۱ - ساخت تصویر جدید از چهره انسان از جمله کاربردهای شبکه GAN است، به‌طور مثال تصویر هیچ‌کدام تصویر واقعی یک انسان نیست و با استفاده از شبکه GAN و یک مجموعه داده تولیدشده است، معماری PGGAN و StyleGAN از جمله معماری‌هایی هستند که در تولید تصاویری با چهره انسان به کار می‌رود. دیتاست CelebA نیز از جمله **دیتاست‌های معروف** مورد استفاده در این حوزه است که ۲۰۲۵۹۹ تصویر چهره را از ۱۰۱۷۷ فرد مشهور جمع‌آوری کرده است. برای آشنایی بیشتر با این دیتاست می‌توانید از این [لینک](#) استفاده نمایید.

برای دانلود مقاله PGGAN رو این [لینک](#) و برای دانلود مقاله StyleGAN رو این [لینک](#) کلیک نمایید. البته این معماری‌ها تنها برای تولید تصاویر انسان به کار نمی‌روند و با دیتاست‌های متفاوت می‌توان تصاویر متفاوتی تولید کرد.



تصویر ۸) نمونه خروجی مقاله PGGAN

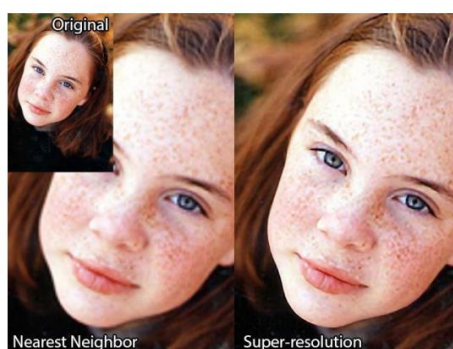
۲ - تبدیل متن به تصویر مثل StackGAN

از کاربردهای خیلی معروف و جذاب GAN می‌توان به تبدیل متن به تصویر اشاره کرد، در این حالت یک متن به شبکه داده می‌شود و متناسب با متن شبکه GAN به شما یک تصویر را به‌عنوان خروجی می‌دهد. به‌طور مثال در تصویر زیر یک متن شامل توصیف یک پرنده را دریافت می‌کند و آن را به تصویر می‌کشد، یکی از معروف‌ترین مدل‌ها در این حوزه مدل StackGAN است که مقاله آن را می‌توانید از [اینجا](#) دانلود نمایید.



تصویر ۹) نمونه تصاویر تولیدشده از متن با استفاده از StackGAN

۳ - افزایش رزولوشن تصویر: یکی دیگر از کاربردهای شبکه GAN افزایش رزولوشن یا به عبارتی سائز تصویر همزمان با حفظ کیفیت تصویر است، همان‌طور که در تصویر زیر مشخص است، روش Nearest Neighbor با روش SRGAN مقایسه شده است که مبتنی بر معماری GAN رزولوشن تصویر را ارتقا می‌دهد، این مقاله را می‌توانید از اینجا دریافت کنید.



تصویر ۱۰) کیفیت افزایش رزولوشن تصویر با SRGA در مقایسه با روش مبتنی بر Nearest Neighbor

۴ - تبدیل تصویر به اموجی از کاربردهای دیگر شبکه GAN است که همان‌طور که در تصویر زیر مشاهده می‌کنید چهره اشخاص را با استفاده از یک مدل GAN به اموجی تبدیل کرده‌اند، در این مقاله نیز از دیتاست CelebA که قبل‌تر معرفی شد استفاده کردند. مقاله را از طریق این [لینک](#) می‌توانید دریافت کنید.

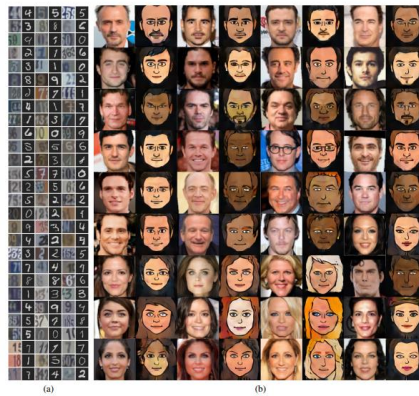
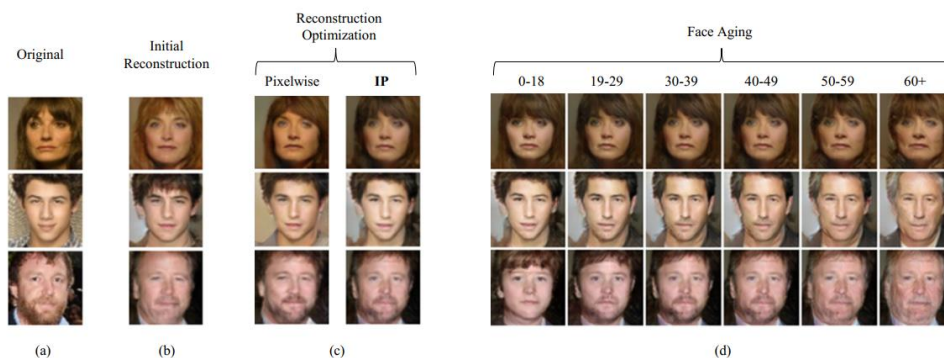


Figure 2: Domain transfer in two visual domains. Input in odd columns; output in even columns. (a) Transfer from SVHN to MNIST. (b) Transfer from face photos (Facescrub dataset) to emoji.

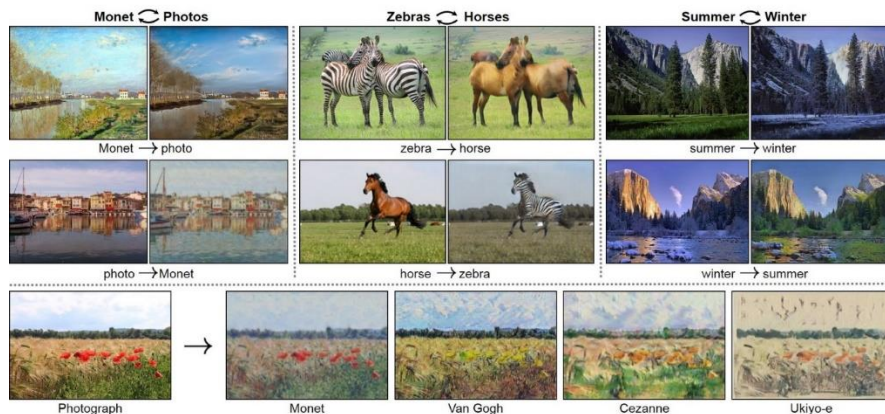
تصویر ۱۱) نمونه‌ای از تبدیل چهره اشخاص به اموجی توسط GAN

۵ - پیر یا جوان کردن افراد از کاربردهای جذاب شبکه‌های GAN است که امروزه آن را در قالب اپلیکیشن‌های مختلف نیز می‌بینیم، یکی از بهترین کارها در این حوزه مربوط به مقاله Face Aging With Conditional Generative Adversarial Networks است که در سال ۲۰۱۸ در کنفرانس CVPR ارائه شده است، این مقاله را می‌توانید از [اینجا](#) دانلود نمایید.



تصویر ۱۲) نمونه‌ای از افزایش یا کاهش سن افراد با استفاده از GAN

۶ - یکی دیگر از کاربردهای شبکه GAN تبدیل تصویر به تصویر است، مثلاً در تصویر زیر می‌بینید که تصویری از محیط از حالت تابستان به زمستان یا برعکس تبدیل شده است، همچنین تصویر گورخری به تصویر اسب و برعکس تبدیل شده است یا تصاویر به نقاشی تبدیل شده‌اند. CycleGAN و Pix2Pix از جمله معماری‌های ارائه شده برای این کار هستند! که مقاله [CycleGAN](#) را از اینجا و مقاله [Pix2Pix](#) را از اینجا می‌توانید دانلود کنید.



تصویر ۱۳) نمونه‌ای از تبدیل تصاویر با استفاده از معماری CycleGAN

سؤالات:

الف) فقط یکی از ۶ کاربرد معرفی شده را انتخاب نمایید و یکی از مقالات معرفی شده مربوط به آن کاربرد را مطالعه نمایید و سازوکار آن شبکه و مدل را توضیح دهید.

ب) مدل انتخابی در قسمت الف را یا با استفاده از یک مدل از پیش آموزش داده شده اجرا کنید یا آن را با استفاده از دیتاست مرتبط به آن که در مقاله معرفی شده است پیاده‌سازی کنید و نتایج را مانند آنچه در صفحه ۴ خواسته شده در گزارش بیاورید. مراحل پیاده‌سازی یا استفاده از Pre-trained model را داخل فایل notebook کد توضیح دهید.

نکته: در دیتاست‌های حجیم فقط قسمتی از دیتاست را برای آموزش شبکه استفاده کنید که آموزش آن‌ها امکان‌پذیر باشد. هدف همان‌طور که در ابتدای تمرین توضیح داده شد یادگرفتن نحوه پیاده‌سازی یا استفاده از این مدل‌ها می‌باشد و نه رسیدن به نتایجی مانند مقالات

نکته: در مورد این سؤال به‌طور کلی پیشنهاد می‌شود مدل‌ها را با استفاده از Pre-trained model اجرا کنید. قاعدتاً دیگر لازم نیست در چنین شرایطی نتایج برای ۵ اپاک متفاوت خروجی را نمایش دهید و فقط کافی است که نحوه عملکرد مدل را توضیح دهید و نمونه تصاویر تولیدشده توسط آن را در گزارش بیاورید.

نکات:

- مهلت تحویل این مینی پروژه، سه‌شنبه ۳۱ تیر است.
 - انجام این مینی پروژه به صورت یک یا دو نفره می‌باشد.
 - گزارش را در قالب تهیه‌شده که روی صفحه درس در CECM بارگذاری شده، بنویسید.
 - گزارش شما در فرآیند تصحیح از اهمیت ویژه‌ای برخوردار است. لطفاً تمامی نکات و فرض‌هایی که برای پیاده‌سازی‌ها و محاسبات خود در نظر می‌گیرید را در گزارش ذکر کنید.
 - در گزارش خود برای تصاویر زیرنویس و برای جداول هم بالانویس اضافه کنید.
 - الزامی به ارائه توضیح جزئیات کد در گزارش نیست؛ اما استثناً در این مینی پروژه باید قسمت‌هایی از کد خود را توضیح دهید که بدین منظور به مطالب خواسته شده در صورت سؤال‌ها توجه نمایید، همچنین نتایج به دست آمده را گزارش و تحلیل کنید.
 - دستیاران آموزشی ملزم به اجرا کردن کدهای شما نیستند؛ بنابراین هرگونه نتیجه و یا تحلیلی که در شرح سؤال از شما خواسته شده است را به‌طور واضح و کامل در گزارش بیاورید. در صورت عدم رعایت این مورد، بدیهی است که از نمره تمرین کسر می‌شود.
 - در صورت مشاهده تقلب نمرات تمامی افراد شرکت‌کننده در آن صفر لحاظ می‌شود.
 - برای انجام تمرین‌ها و مینی پروژه‌ها فقط مجاز به استفاده از زبان برنامه‌نویسی Python خواهید بود.
 - استفاده از کدهای آماده برای تمرین‌ها به هیچ‌وجه مجاز نیست؛ اما برای مینی پروژه‌ها فقط برای قسمت‌هایی از کد و به‌عنوان راهنمایی برای پیاده‌سازی، می‌توانید از کدهای آماده استفاده کنید.
 - در این مینی پروژه هیچ محدودیتی از نظر کتابخانه‌های مورد استفاده در زبان پایتون ندارید.
 - نحوه محاسبه تأخیر به این شکل است: مهلت بدون کسر نمره تا تاریخ اعلام‌شده و پس از آن به مدت هفت روز تا ۷ مرداد با ۳۰ درصد کسر نمره، بارگذاری ممکن است و درنهایت، پس از بازه تأخیر نمره تمرین صفر خواهد شد.
 - لطفاً گزارش، فایل کدها و سایر ضمایم موردنیاز را با فرمت زیر در سامانه مدیریت دروس بارگذاری نمایید (ارسال و ذکر نام یکی از اعضای گروه در نام فایل کفایت می‌کند).
- Project3_[Lastname]_[StudentNumber].zip
- در صورت وجود هرگونه ابهام یا مشکل می‌توانید از طریق رایانامه‌های زیر با دستیاران آموزشی مربوطه در تماس باشید:

Dehghanr.mohammad@gmail.com

Alikarimi120@gmail.com

پیوست:

نمونه تصاویر MNIST:

3	9	8	3	7	0	3	1	1	1
8	5	3	9	6	4	0	7	0	6
9	9	0	9	6	3	1	7	5	4
2	1	9	3	0	6	0	5	9	8
2	6	8	9	9	4	4	8	9	1
1	5	0	3	3	8	4	2	7	6
8	2	0	4	7	8	1	8	6	6
2	8	7	0	0	9	8	3	6	5
0	7	0	5	5	4	0	1	4	9
9	5	1	4	0	8	3	4	8	9

نمونه تصاویر CIFAR-10:

