



پروژه درس رایانش عصبی و یادگیری عمیق

پروژه‌ی ششم

هدف: بررسی کارایی شبکه‌های مولد تبدیل متن به تصویر

کد: پیاده سازی این پروژه را به زبان پایتون انجام دهید؛ در این فعالیت مجاز به استفاده از tensorflow یا pytorch یا jax می‌باشید. فایل‌های کد خود را بر اساس شماره سوال و زیر قسمت خواسته شده‌ی آن نام گذاری کنید (برای مثال می‌توان نام گذاری قسمت اول برای سوال سوم تمرین را بصورت P3_a_preprocessing.py در نظر گرفت). فایل‌های ارسالی‌تان بایستی با فرمت py یا ipynb (با حفظ خروجی هر سلول) باشد.

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 که دارای فهرست بوده و پاسخ‌ها بترتیب در آن قرار گرفته اند و نام، نام خانوادگی و شماره دانشجویی‌تان در قسمت چپ سربرگ تمامی صفحات تکرار شده است. علاوه بر خواسته‌ی مستقیم هر سوال، مقتضی است که نمودارهای خطا (loss) و صحت (accuracy) را به ازای مجموعه داده‌های آموزش و اعتبارسنجی رسم نمایید. همچنین در صورت امکان ماتریس درهم‌ریختگی را بصورت رنگ‌آمیزی شده به همراه اعداد متناظر برای مجموعه داده‌های آموزش، آزمون و اعتبارسنجی نیز تولید نمایید. لازم به ذکر است که در هر آموزش بایستی موارد مهم تنظیم شده نظیر تابع خطا، بهینه‌ساز (به همراه پارامترهای تنظیم شده‌ی آن مانند نرخ یادگیری)، معماری شبکه‌ی آموزشی (کتابخانه‌ها و ابزارهایی برای بصری‌سازی موجود است)، تعداد گام آموزشی، اندازه دسته (Batch Size)، آمارگان تفکیک مجموعه داده (به آموزش، آزمون و اعتبارسنجی)، پیش‌پردازش‌های اعمالی بروی دادگان ورودی و... ذکر گردد.

تذکر: مطابق قوانین دانشگاه هر نوع کپی برداری و اشتراک کار دانشجویان غیر مجاز بوده و با تمامی طرفین برخورد خواهد شد. استفاده از کدها و توضیحات اینترنت به منظور یادگیری صرفاً با ارجاع به آن بلامانع است، اما کپی کردن آن غیرمجاز است.

راهنمایی: در صورت نیاز می‌توانید سوالات خود را در خصوص پروژه از تدریس‌یارهای درس، از طریق ایمیل زیر یا در گروه تلگرامی بپرسید. ([لینک گروه تلگرامی](#))

Email: ann.ceit.aut@gmail.com CC: m.ebadpour@aut.ac.ir

توجه: می‌توانید از منابع و بسترهای سخت افزاری برخط رایگان نظیر Google Colab یا Kaggle استفاده نمایید.

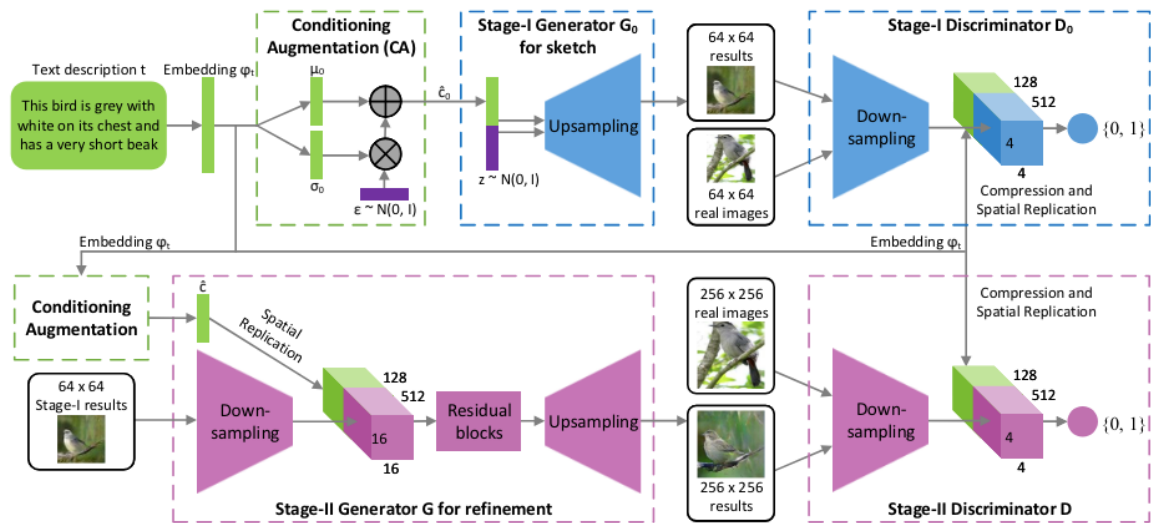
تاخیر مجاز: در طول ترم، ده روز زمان مجاز تاخیر برای ارسال پروژه‌ها در اختیار دارید (بدون کسر نمره). این تاخیر را می‌توانید بر حسب نیاز بین پروژه‌های مختلف تقسیم کنید که مجموع آن نباید بیشتر از ده روز شود. پس از استفاده از این تاخیر مجاز، هر روز تاخیر باعث کسر ۱۰٪ نمره‌ی کسب شده‌ی آن تمرین خواهد شد.

ارسال: فایل های کد و گزارش خود را در قالب یک فایل فشرده با فرمت StudentID_HW06.zip تا تاریخ ۱۴۰۳/۰۴/۱۴ صرفاً از طریق سایت کورسز ارسال نمایید. ارسال از طریق تلگرام، ایمیل و سایر راههای ارتباطی مجاز نبوده و تصحیح صورت نخواهد گرفت.

قسمت اول: شبکه‌های مولد تقابلی

شبکه‌های مولد تقابلی^۱ همانطور که در کلاس با آنها آشنا شدید شامل دو زیرشبکه‌ی تولیدکننده^۲ و تمایزگر^۳ هستند که به صورت تقابلی آموزش داده می‌شوند تا داده‌های جدید تولید کنند. تولید داده‌ی جدید هدفی است که در تمامی مدل‌های مولد مد نظر قرار دارد و به شکل‌های مختلف از جمله ترجمه‌ی تصویر به تصویر، تبدیل دامنه و تولید شرطی صورت می‌گیرد. یکی از این اشکال، تولید تصویر با دریافت فرمان زبانی است که امروزه نیز نمونه‌های کاربردی آن همچون Dall-E و Imagen در دسترس عموم قرار دارند. در این تمرین به طور خاص به پیاده سازی این وظیفه با شبکه‌ی مولد تقابلی پشته‌ای یا SatckGAN می‌پردازیم.

۱. با مراجعه به [مقاله‌ی StackGAN](#) کلیت ساختار و چگونگی عملکرد این شبکه را توضیح دهید. توضیح دهید که شبکه‌ی تعریف شده در هر گام^۴ به چه منظور استفاده می‌شود. به طور خاص ذکر کنید که ورودی شبکه‌ی تولیدکننده در هر دو گام چه تفاوتی با ورودی یک شبکه‌ی مولد تقابلی ساده^۵ دارد؟ همچنین بررسی کنید که آموزش این شبکه به چه صورت انجام می‌شود. (۱۰ امتیاز)



تصویر ۱: معماری کلی شبکه‌ی مولد تقابلی پشته‌ای

¹ Generative Adversarial Networks

² Generator

³ Discriminator

⁴ Stage




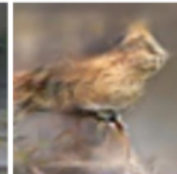




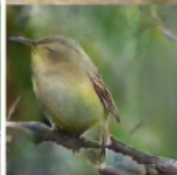

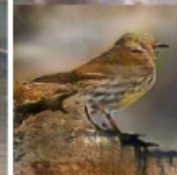



⁵ Vanilla GAN

2 . شبکه‌های مولد تقابلی در مقایسه با سایر شبکه‌ها از سه مشکل اساسی رنج می‌برند؛ این سه مشکل عبارتند از فروپاشی مد⁶، عدم همگرایی و ناپدید شدن گرادیان. به طور مختصر توضیح دهید که هر کدام به چه صورتند و چه راهکارهایی برای رفع آنها مطرح شده است؟(۵ امتیاز)

3 . یک ایده‌ی رایج برای بهبود عملکرد شبکه‌های مولد تقابلی استفاده از عملگر PixelShuffle است. نحوه‌ی عملکرد این عملگر و تاثیر آن را بررسی کنید. بررسی کنید که این عملگر اولین بار در چه وظیفه‌ای و به چه منظور تعریف شد؟ همچنین بررسی کنید که به طور خاص در معماری StackGAN در کدام زیرشبکه‌ها قابل استفاده است و چه عملکردی خواهد داشت؟(۷ امتیاز)

4 . معیار FID(Frechet Inception Score) یک معیار برای ارزیابی کیفیت و تنوع تصاویر تولید شده توسط مدل‌های مولد است. توضیح دهید که این معیار به چه صورت محاسبه می‌شود، به چه ویژگی‌هایی از مدل و یا داده وابسته است و آیا معیار قابل اتکایی برای مقایسه‌ی مدل‌های مولد محسوب می‌شود؟(۸ امتیاز)

برای این پروژه از مجموعه داده‌ی CUB⁻2011 استفاده می‌کنیم که شامل یازده هزار تصویر از ۲۰۰ گونه پرنده می‌باشد و به ازای هر تصویر یک توصیف متنی نیز وجود دارد. [مجموعه داده](#) در سایت Kaggle و توصیفات متنی نیز در [این لینک](#) موجودند. همچنین برای توصیفات متنی نیز یک تعبیه‌ی از پیش آماده شده در فایل char⁻CNN⁻RNN⁻embeddings.pickle وجود دارد که می‌تواند جایگزین ساخت تعبیه از مدل‌های از پیش آموزش داده شده باشد. استفاده از سایر تعبیه‌ها نیز که منجر به کارایی بهتر مدل شوند دارای ۵ امتیاز اضافی می‌باشد.

Text description	This bird is blue with white and has a very short beak	This bird has wings that are brown and has a yellow belly	A white bird with a black crown and yellow beak	This bird is white, black, and brown in color, with a brown beak	The bird has small beak, with reddish brown crown and gray belly	This is a small, black bird with a white breast and white on the wingbars.	This bird is white black and yellow in color, with a short black beak
Stage-I images							
Stage-II images							

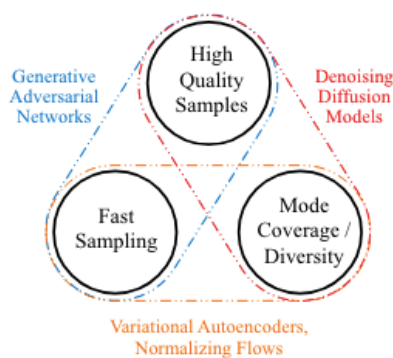
تصویر ۲: نمونه‌ی خروجی مدل StackGAN برای مجموعه داده‌ی CUB

⁶ Mode Collapse

5. مدل را بر روی این داده‌ها آموزش دهید. معماری نهایی هر یک از چهار زیر شبکه به همراه نمودار خطای تولید کننده و تمایزگر در هر گام آموزش را در گزارش خود بیاورید. پس از پایان آموزش ۱۰ تصویر را به صورت تصادفی از خروجی مدل در stage اول و دوم تولید کنید. (۵۰ امتیاز)

قسمت دوم: مدل‌های پخشی⁷ (بخش امتیازی)

مدل‌های مولد حوزه‌ی تصویر به چهارچوب‌های مختلف تقسیم می‌شوند. وجه مشترک همه‌ی این مدل‌ها این است که تلاش می‌کنند تا با یادگیری توزیع داده‌ها نمونه‌های جدیدی از آن تولید کنند. تاکنون مدل‌های مولد حوزه‌ی تصویر را می‌توان در چهار قالب کلی شامل خودکدگذارهای تغییراتی⁸، شبکه‌های مولد تقابلی، جریان‌های نرمال‌ساز⁹ و مدل‌های پخشی دسته‌بندی کرد که در هر قالب انواع مختلفی از پیاده‌سازی‌ها وجود دارد.



تصویر ۳: مشکل سه‌گانه‌ی مدل‌های مولد

1. در [این مقاله](#) سه نیازمندی کلی برای کارایی یک مدل مولد حوزه‌ی تصویر ذکر می‌شود که عبارتند از: تولید نمونه‌های با کیفیت، سرعت بالای تولید نمونه و تنوع نمونه‌های تولیدی. و نیز اشاره می‌شود که هر مدل مولدی که تاکنون در یکی از قالب‌هایی که بالاتر ذکر شد ارائه شده است در یکی از این سه نیازمندی ضعیف عمل می‌کند. با بررسی مقاله توضیح دهید که هر مدل در چه زمینه‌ای و به چه علتی ضعیف عمل می‌کند؟ (۵ امتیاز)

مدل‌های پخشی دسته‌ای از مدل‌های مولد هستند که در حال حاضر به عنوان بهترین مدل تولید تصویر شناخته می‌شوند. در مدل‌های پخشی احتمالاتی از یک زنجیره‌ی مارکف برای مدل کردن فرآیند نویززدایی و نویز افزایی استفاده می‌شود و دو مسیر کلی رو به جلو¹⁰ و رو به عقب¹¹ در نظر گرفته می‌شود. در مسیر روبه جلو داده‌ی اولیه مرحله به مرحله با نویز تخریب می‌شود تا به یک نویز تماماً گاوسی تبدیل شود و در فرآیند رو به عقب نیز نویز زدایی با شروع از یک نویز تصادفی اولیه انجام می‌شود تا به نمونه‌ای جدید از توزیع داده‌ها برسیم. این موضوع در تصویر ۴ آورده شده است.

⁷ Diffusion Models

⁸ Variational Autoencoders

⁹ Normalizing Flows

¹⁰ Forward

¹¹ Backward



تصویر ۴: فرآیند کلی مدل‌های پخش

2. با توجه به [مقاله‌ی مدل‌های پخش احتمالاتی](#)، در مسیر رو به جلو نیازی به اضافه کردن نویز به صورت مرحله به مرحله نیست و می‌توان نویز اضافه شونده به هر مرحله را به صورت مستقیم و با استفاده از رابطه‌ی زیر به دست آورد:

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\alpha_t}x_0, (1 - \alpha_t)I)$$

با استفاده از خاصیت نویز گاوسی رابطه‌ی بالا را اثبات کنید. ([راهنمایی](#)): اگر دو متغیر نرمال مستقل داشته باشیم جمع آنها نیز نرمال است. (۱۰ امتیاز)

3. با توجه به مقاله‌ی سوال ۲، فرآیند آموزش و نمونه برداری مدل‌های پخش را توضیح دهید. در فرآیند رویه عقب یک فرض مهم این است که توزیع $q(x_{t-1}|x_t)$ گاوسی است؛ در چه صورتی این فرض درست است؟ (۵ امتیاز)

4. یک مساله‌ی اساسی در مدل‌های پخش این است که در هیچ یک از گام‌ها محاسبات در بعد کوچکتري صورت نمی‌گیرند در نتیجه در صورت بزرگ بودن دیتاست و اندازه‌ی تصاویر ورودی مدل‌های پخش بسیار پرهزینه و حجیم خواهد شد. مدل latent stable diffusion برای حل این چالش از چه رویکردی استفاده می‌کند؟ به نظر شما چرا برای کاهش حجم نیاز به مدل‌های تغییراتی¹² داریم؟ (۵ امتیاز)

5. یک مدل پخش را بر روی داده‌های MNIST آموزش دهید. تعداد پارامترهای مدل، تصاویر میانی فرآیند رو به جلو و فرآیند رو به عقب را برای یک تصویر گزارش کنید. همچنین ۵ تصویر تولید شده توسط این مدل را نمایش دهید. (۱۵ امتیاز)