

شبکه‌های عصبی و یادگیری عمیق

دکتر صفا بخش



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)
دانشکده مهندسی کامپیوتر

رضا آدینه پور ۴۰۲۱۳۱۰۵۵

تمرین ششم
شبکه GAN

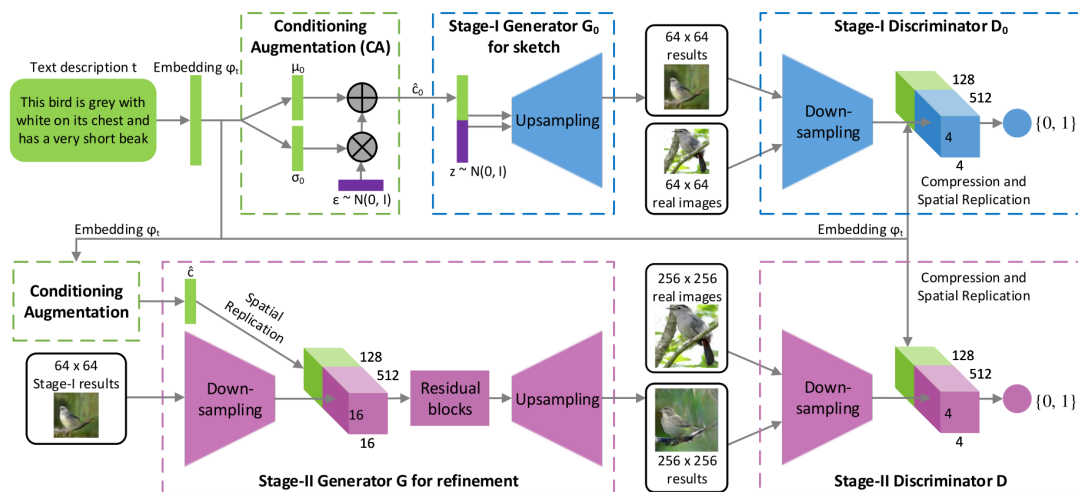
۱۴ تیر ۱۴۰۳



سوال 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 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 Imagen و Dall-E در دسترس عموم قرار دارند. در این تمرین به طور خاص به پیاده سازی این وظیفه با شبکه‌ی مولد تقابلی پشته‌ای یا SatchGAN می‌پردازیم.

۱. به مراجعه به مقاله **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.



شکل ۱: معماری کلی شبکه مولد تقابلی پشته‌ای

پاسخ

NetworksAdversarial Generative^۱
 Generator^۲
 Discriminator^۳
 Stage^۴
 Vanilla GAN^۵

۲. شبکه‌های مولد تقابلی در مقایسه با سایر شبکه‌ها از سه مشکل اساسی رنج می‌برند: این سه مشکل عبارتند از فروپاشی مد، عدم همگرایی و ناپدید شدن گرادیان. به طور مختصر توضیح دهید که هر کدام به چه صورتی و چه راهکارهایی برای رفع آنها مطرح شده است؟

پاسخ

۳. یک ایده‌ی رایج برای بهبود عملکرد شبکه‌های مولد تقابلی استفاده از عملکرد PixelShuffle است. نحوه‌ی عملکرد این عملکرد و تأثیر آن را بررسی کنید. بررسی کنید که این عملکرد 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؟

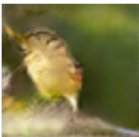

پاسخ

۴. 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؟

پاسخ

۵. مدل را بر روی این داده‌ها آموزش دهید. معماری نهایی هر یک از چهار زیرشبکه به همراه نمودار خطای تولیدکننده 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.

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. **مجموعه‌ی داده** در سایت Kaggle و توصیفات متنی نیز در **این لینک** موجود است. همچنین برای توصیفات متنی از پیش تعبیه‌ی آماده شده در فایل char-CNN-RNN-embeddings.pickle وجود دارد که می‌تواند جایگزین ساختن داده باشد. استفاده از پیش 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							

شکل ۲: نمونه خروجی مدل StackGAN برای مجموعه داده CUB

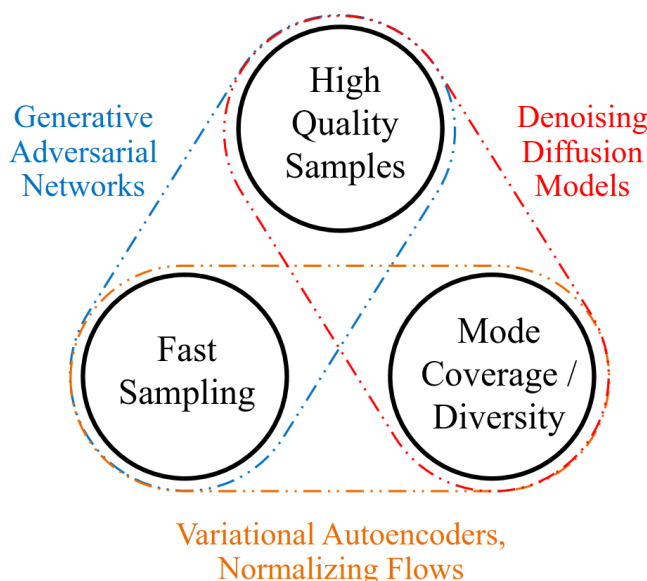
Mode Collapse

پاسخ

سوال دوم - مدل‌های پخشی (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 تغییراتی^۷، شبکه‌های مولد تقابلی، جریان‌های نرمال‌ساز^۸ و مدل‌های پخشی دسته‌بندی کرد که در هر قالب انواع مختلفی از پیاده‌سازی‌ها وجود دارد.

۱. 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؟



شکل ۳: مشکل سه‌گانه مدل‌های مولد

مدل‌های پخشی دسته‌ای از مدل‌های مولد هستند که در حال حاضر بهترین مدل تولید تصویر شناخته می‌شوند. در مدل‌های پخشی احتمالاتی از یک ذخیره‌ی مارکف برای مدل کردن فرآیند نویززدایی و نویز افزایشی استفاده می‌شود و دو مسیر کلی رو به جلو^۹ و رو به عقب^{۱۰} در نظر گرفته می‌شود. در مسیر رو به جلو داده‌ی اولیه مرحله به مرحله با نویز تخریب می‌شود تا به یک نویز 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.



شکل ۴: فرایند کلی مدل‌های پخشی

Diffusion Models^۷
Flows Normalizing^۸
Forward^۹
Backward^{۱۰}

۱. با توجه به مقاله‌ی **مدل‌های پخشی 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:

$$q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1 - \bar{\alpha}_t)\mathbf{I})$$

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

۲. با توجه به مقاله‌ی سوال قبل، فرآیند آموزش و نمونه برداری مدل‌های پخشی را توضیح دهید. در فرآیند رو به عقب یک فرض مهم این است که توزیع $q(\mathbf{x}_{t-1}|\mathbf{x}_t)$ گاوسی است. در چه صورتی این فرض درست است؟

۳. یک مسالهی اساسی در مدل‌های پخشی 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 پخشی 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 تغییراتی^{۱۱} داریم؟

۴. یک مدل پخشی را بر روی داده‌های MNIST آموزش دهید. تعداد پارامترهای مدل، تصاویر میان 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.