

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

فهرست مطالب

فصل 1: مقدمه	5
1-1 شبکه‌های متخاصم مولد (GAN) چیست؟	5
1-2 تاریخچه مختصر GAN ها	5
1-3 ایده اصلی: آموزش خصمانه	6
1-4 کاربردهای GAN	8
فصل 2: شبکه GAM	9
2-1 شبکه مولد	9
2-2 شبکه متمایز کننده	10
2-3 تابع ضرر	13
2-4 فرآیند آموزش	15
فصل 3: انواع GAN	18
3-1 شبکه گن شرطی	18
2-3 GAN های کانولوشنال عمیق	20
3-3 شبکه Wasserstein GANs (WGANs) [4]	22
3-4 شبکه‌های متخاصم مولد چرخه	23
5-3 شبکه StyleGAN	26
6-3 شبکه Progressive GAN [7]	28
3-7 شبکه متخاصم سازنده متوالی	30
فصل 4: چالش‌ها و محدودیت‌ها	33
4-1 بی‌ثباتی آموزش	33
4-1-1 فروپاشی حالت	33
4-1-2 ناپدید شدن گرادیان	34
4-1-3 جمع‌بندی	34
فصل 5: معیارهای ارزیابی	36
5-1 امتیاز اولیه (IS): اندازه‌گیری کیفیت و تنوع	36
5-2 فاصله شروع فریشت (FID): اندازه‌گیری قوی‌تر	37

38	5-3 الزامات داده و هزینه محاسباتی: قیمت واقع گرایی
38	5-3-1 نیازهای داده
38	5-3-2 هزینه محاسباتی: تلاشی با منابع فشرده
39	5-3-3 کاهش چالش‌ها
39	5-4 نگرانی‌های اخلاقی و سوء استفاده بالقوه: سمت تاریک خلقت
39	5-4-1 دیپ فیک و اطلاعات غلط: فرسایش اعتماد
40	5-4-2 نگرانی‌های حفظ حریم خصوصی: محو شدن مرزها
40	5-4-3 نقض حق چاپ: مسئله مالکیت
40	5-4-4 تقویت بایاس: اتاق پژواک هوش مصنوعی
40	5-4-5 نیاز به توسعه و استقرار مسئولانه
42	فصل 6: کاربردهای GAN در جزئیات
42	6-1 تولید و دستکاری تصویر: یک انفجار خلاقانه
44	فصل 7: جهت‌گیری‌های آینده و روندهای تحقیقاتی
44	7-1 بهبود ثبات آموزش: تلاش برای استحکام
44	7-2 افزایش کیفیت و تنوع تصویر: جستجوی کمال
45	7-3 کاوش در معماری‌های جدید و عملکردهای زیان: موتور نوآوری
45	7-4 پرداختن به نگرانی‌های اخلاقی: ضرورت هوش مصنوعی مسئول
47	فصل 8: نتیجه‌گیری
49	فصل 9: منابع

فهرست شکل‌ها

- شکل 1-1 کارکرد کلی شبکه *gan* 7
- شکل 2-1 شبکه مولد 9
- شکل 2-2 تشخیص داده واقعی از ساختگی 11
- شکل 2-3 خروجی شبکه متمایزکننده 12
- شکل 2-4 ناسازگاری‌های به وجود آمده در زمان آموزش 12
- شکل 3-1 نوع ساختار *conditional gan* 18
- شکل 3-2 ساختار شبکه *drgan* 20
- شکل 3-3 نمونه تصویر ساخته شده توسط *cyclegan* 24
- شکل 3-4 ساختار شبکه *cyclegan* 25
- شکل 3-5 چند نمونه تصویر ساخته شده توسط *stylegan* 27
- شکل 3-6 ساختار شبکه *stylegan* 27
- شکل 3-7 ساختار شبکه *seq gan* 30
- شکل 4-1 نمونه تصویر ساخته شده توسط شبکه در فروپاشی حالت 33

فصل 1: مقدمه

1-1 شبکه‌های متخاصم مولد (GAN) چیست؟

شبکه‌های متخاصم مولد¹ (GAN) دسته‌ای از مدل‌های یادگیری ماشین هستند که برای یادگیری توزیع احتمال اساسی² یک مجموعه داده و تولید نمونه‌های داده جدید با ویژگی‌های مشابه طراحی شده‌اند. برخلاف مدل‌های مولد سنتی که بر تخمین چگالی احتمال صریح تکیه می‌کنند، GANها از یک رویکرد «متخاصم»³ هوشمندانه استفاده می‌کنند و دو شبکه عصبی را در برابر یکدیگر قرار می‌دهند: یک مولد و یک تمایز.

1-2 تاریخچه مختصر GANها

داستان GANها با ایان گودفلو⁴ و همکاران در دانشگاه مونترال آغاز می‌شود که آن‌ها را در مقاله مهم خود در سال 2014 با نام "شبکه‌های متخاصم مولد [1]" معرفی کردند. این مقاله پایه و اساس یک پارادایم جدید در مدل‌سازی مولد را ایجاد کرد و از روش‌های سنتی که بر تعریف صریح توابع چگالی احتمال تکیه داشتند، فاصله گرفت. نوآوری اصلی فرآیند آموزش خصمانه بود که از تئوری بازی⁵ الهام گرفته شده بود، جایی که دو شبکه عصبی، مولد⁶ و متمایزکننده⁷، با یکدیگر رقابت می‌کنند. این رویکرد جدید به GANها اجازه می‌دهد تا توزیع‌های داده‌های پیچیده را بدون نیاز به فرمول‌های پیچیده ریاضی آن توزیع‌ها یاد بگیرند.

استقبال اولیه از GANها مشتاقانه بود، اگرچه فوراً انفجاری نبود. در حالی که پتانسیل روشن بود، آموزش GANها چالش برانگیز بود، و نتایج اولیه، در عین چشمگیر بودن، برخی محدودیت‌ها را نیز برجسته کرد. با این حال، جامعه یادگیری ماشینی به سرعت اهمیت این تکنیک جدید، به ویژه برای تولید تصویر را تشخیص داد. توانایی GANها برای تولید تصاویر کاملاً واقعی، موجی از تحقیق و توسعه را برانگیخت.

پس از انتشار اولیه، حوزه GANها رشد سریعی را تجربه کرد. محققان معماری‌های مختلف شبکه، توابع خطا و استراتژی‌های آموزشی را برای بهبود پایداری و عملکرد GANها بررسی کردند. پیشرفت‌های کلیدی شامل توسعه Deep Convolutional GANs (DCGANs) بود که از قدرت شبکه‌های عصبی کانولوشن برای تولید تصویر استفاده کرد و کیفیت و وضوح تصاویر تولید شده را به طور قابل توجهی بهبود بخشید. این پیشرفت GANها را برای کاربردهای دنیای واقعی کاربردی‌تر کرد.

¹Generative Adversarial Networks
²underlying probability distribution
³adversarial
⁴Ian Goodfellow
⁵Game theory
⁶Generator
⁷Discriminator

با پیشرفت تحقیقات، تغییرات و توسعه‌های متعددی از GANها پدیدار شد که هر کدام محدودیت‌های خاصی را مورد توجه قرار می‌دهند یا برنامه‌های کاربردی جدید را هدف قرار می‌دهند. GANهای شرطی¹ (CGANs) با ترکیب اطلاعات اضافی مانند برچسب‌های کلاس، تولید کنترل‌شده را فعال می‌کنند و امکان سنتز هدفمند تصویر را فراهم می‌کنند. Wasserstein GANs (WGANs) با معرفی یک تابع ضرر جدید بر اساس فاصله Wasserstein که منجر به آموزش پایدارتر و قابل اعتمادتر می‌شود، با موضوع بی‌ثباتی در زمان آموزش مقابله کرد. CycleGANها رویکرد جدیدی را برای ترجمه (translate) تصویر به تصویر ارائه کردند که امکان تبدیل تصاویر از یک دامنه به دامنه دیگر را بدون نیاز به داده‌های آموزشی جفتی فراهم می‌کرد. StyleGAN که توسط انویدیا توسعه داده شد، انقلابی در زمینه سنتز² تصویر ایجاد کرد و تصاویری با وضوح فوق‌العاده بالا و واقعی را به همراه کنترل دقیق بر محتوای تولید شده تولید کرد.

تکامل GANها با جستجوی مداوم برای معماری‌های بهتر، توابع خطا و تکنیک‌های آموزشی مشخص شده است. محققان راه‌های جدیدی را برای مقابله با چالش‌هایی مانند فروپاشی حالت³، ناپدید شدن گرادیان‌ها⁴ و معیارهای ارزیابی کشف کرده‌اند. این زمینه با تحقیقات مداوم برای بهبود کیفیت، تنوع و کنترل پذیری داده‌های تولید شده و همچنین گسترش کاربردهای GANها در حوزه‌های جدید بسیار فعال است. از تولید تصاویر و ویدیوهای واقعی گرفته تا افزایش وضوح تصویر، ترجمه بین حوزه‌های تصویر و حتی کشف داروهای جدید، GANها پتانسیل خود را برای ایجاد انقلاب در زمینه‌های مختلف نشان داده‌اند.

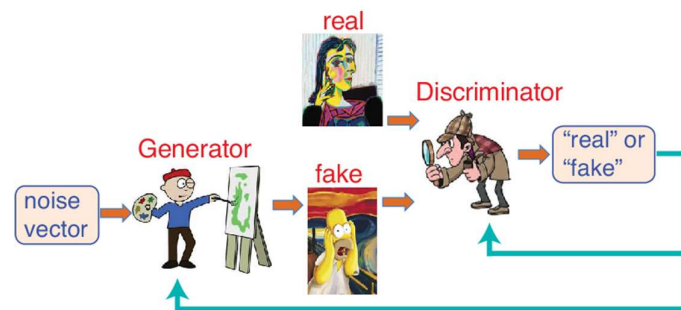
3-1 ایده اصلی: آموزش خصمانه

مفهوم اصلی پشت GANها آموزش خصمانه است، فرآیندی که در آن دو شبکه عصبی، مولد و متمایزکننده، به طور همزمان در یک سناریوی رقابتی و بازی مانند آموزش داده می‌شوند. این رویکرد خصمانه چیزی است که GANها را از سایر مدل‌های مولد متمایز می‌کند. این مانند یک بازی موش و گربه است که در آن مولد تلاش می‌کند تا داده‌های "جعلی" متقاعدکننده را به طور فزاینده‌ای ایجاد کند و متمایزکننده سعی می‌کند این داده‌های جعلی را از داده‌های واقعی تشخیص دهد.

مولد را به عنوان هنرمندی تصور کنید که تلاش می‌کند یک نقاشی واقع‌گرایانه خلق کند، و متمایزکننده را به عنوان یک منتقد هنری که سعی در کشف جعلیات دارد. هنرمند با خط خطی‌های تصادفی (نویز) شروع می‌کند و به تدریج تکنیک خود را اصلاح می‌کند و یاد می‌گیرد که نقاشی‌های بیشتر و واقعی‌تری خلق کند. در همین حال، منتقد هنری، نقاشی‌های واقعی را مطالعه می‌کند و یاد می‌گیرد که سرنخ‌های ظریفی را که آن‌ها را از جعلیات متمایز

¹ Conditional
Image Synthesis²
Mode collapse³
vanishing gradients⁴

می‌کند، شناسایی کند. هر چه هنرمند بهتر می‌شود، منتقد باید بصیرتر شود و هر چه منتقد بهتر می‌شود، هنرمند باید خلاق‌تر شود.



شکل 1-1 کارکرد کلی شبکه gan^1

در چارچوب GAN، شبکه مولد نویز تصادفی را به عنوان ورودی می‌گیرد. این نویز، که اغلب بردار اعداد تصادفی است، به عنوان نقطه شروعی برای فرآیند تولید عمل می‌کند. مولد این نویز را از طریق یک سری لایه تبدیل می‌کند و به تدریج آن را به یک نمونه داده مصنوعی مانند یک تصویر، یک قطعه موسیقی یا یک متن متنی تبدیل می‌کند. از سوی دیگر، شبکه متمایزکننده با دو نوع داده ارائه می‌شود، نمونه‌های داده واقعی برگرفته از مجموعه داده آموزشی و نمونه‌های داده جعلی تولید شده توسط مولد، وظیفه آن طبقه بندی هر ورودی به عنوان واقعی یا جعلی است.

کلید آموزش GAN ماهیت رقابتی این تنظیم است. هدف مولد این است که متمایزکننده را "فریب دهد"²، تا داده‌های مصنوعی ایجاد کند که متمایزکننده به عنوان واقعی طبقه بندی کند. هدف تمایز کننده برعکس است: شناسایی صحیح داده‌های جعلی تولید شده توسط مولد. این دو شبکه، به طور همزمان، با توابع خطا همراه با هم آموزش داده می‌شوند. همانطور که مولد در ایجاد داده‌های واقعی بهتر می‌شود، متمایزکننده مجبور می‌شود انتخابی‌تر شود و توانایی خود را برای تشخیص تقلبی بهبود بخشد. این به نوبه خود، مولد را برای ایجاد داده‌های واقعی‌تر سوق می‌دهد.

این فعل و انفعال پویا هر دو شبکه را به بهبود مکرر سوق می‌دهد. مولد یاد می‌گیرد که الگوها و ساختار زیربنایی توزیع داده‌های واقعی را به تصویر بکشد، و متمایزکننده یاد می‌گیرد که نشانه‌های ظریفی را که داده‌های واقعی را از جعلی متمایز می‌کند، شناسایی کند. از طریق این فرآیند آموزش خصمانه، مولد در نهایت یاد می‌گیرد که داده‌های مصنوعی با کیفیت بالا تولید کند که عملاً از داده‌های واقعی قابل تشخیص نیستند. این یک تکنیک قدرتمند است که منجر به پیشرفت‌های قابل توجهی در مدل سازی تولیدی، به ویژه در زمینه تولید تصویر شده است.

¹https://www.researchgate.net/figure/Generative-adversarial-network-with-the-generator-creating-a-fake-image-eg-a-fake_fig3_359955646

fool²

1-4 کاربردهای GAN

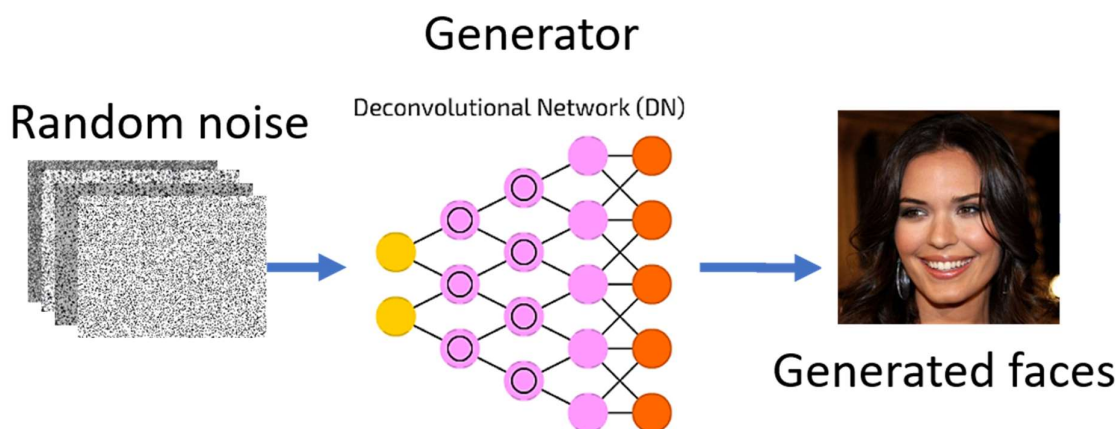
- GANها طیف وسیعی از کاربردها را در دامنه‌های مختلف پیدا کرده‌اند. برخی از نمونه‌های برجسته عبارتند از:
- تولید و دستکاری تصویر: ایجاد تصاویر واقعی از اشیا، افراد و صحنه‌ها و همچنین اصلاح تصاویر موجود.
 - ترجمه تصویر به تصویر: تبدیل تصاویر از یک دامنه به دامنه دیگر (به عنوان مثال، طرح‌ها به عکس، روز به شب).
 - Super-Resolution و Image Enhancement: بهبود وضوح و کیفیت تصاویر.
 - افزایش داده‌ها: تولید داده‌های مصنوعی برای تقویت مجموعه داده‌های آموزشی و بهبود استحکام مدل.
 - تشخیص ناهنجاری: شناسایی الگوهای غیر معمول یا نقاط پرت در داده‌ها.

فصل 2: شبکه GAM

2-1 شبکه مولد

شبکه مولد که نماد آن G است، هنرمند دنیای GAN است. وظیفه آن گرفتن نویز تصادفی و تبدیل آن به داده‌های مصنوعی است که از داده‌های واقعی که روی آن آموزش می‌بیند تقلید می‌کند. مجسمه‌سازی را در نظر بگیرید که با یک بلوک گلی (نویز) شروع می‌کند و آن را به مجسمه‌ای واقعی تبدیل می‌کند (داده‌های مصنوعی). شبکه مولد از یک سری عملیات ریاضی در یک شبکه عصبی برای دستیابی به این تبدیل استفاده می‌کند.

ورودی شبکه مولد معمولاً بردار نویز تصادفی z است. این بردار اغلب از یک توزیع احتمال ساده، معمولاً یک توزیع نرمال استاندارد (میانگین 0، انحراف معیار 1) ترسیم می‌شود. این بردار نویز بسیار مهم است زیرا تصادفی اولیه مورد نیاز برای شبکه مولد برای تولید خروجی‌های متنوع را فراهم می‌کند. اگر ورودی همیشه یکسان بود، مولد همیشه همان خروجی را تولید می‌کرد و توانایی آن را برای ایجاد داده‌های متنوع و جالب محدود می‌کرد. نویز مانند یک دانه عمل می‌کند و بردارهای نویز مختلف منجر به خروجی‌های متفاوت تولید می‌شوند. ابعاد بردار نویز می‌تواند بر پیچیدگی داده‌های تولید شده تأثیر بگذارد. یک بردار نویز با ابعاد بالاتر ممکن است خروجی‌های پیچیده‌تر و دقیق‌تری را ارائه دهد.



شکل 2-1 شبکه مولد

این بردار نویز z وارد شبکه مولد می‌شود که از چندین لایه نورون‌های به هم پیوسته تشکیل شده است. هر لایه یک سری تبدیل ریاضی انجام می‌دهد. اول، عملیات خطی وجود دارد که شامل ضرب بردار ورودی در ماتریسی از وزن‌ها و اضافه کردن یک بردار بایاس است. این وزن‌ها و سوگیری‌ها پارامترهای قابل یادگیری مولد هستند که با θ_g نشان داده می‌شوند. آن‌ها چیزی هستند که مولد در طول فرآیند آموزش می‌آموزد. این عملیات خطی توسط

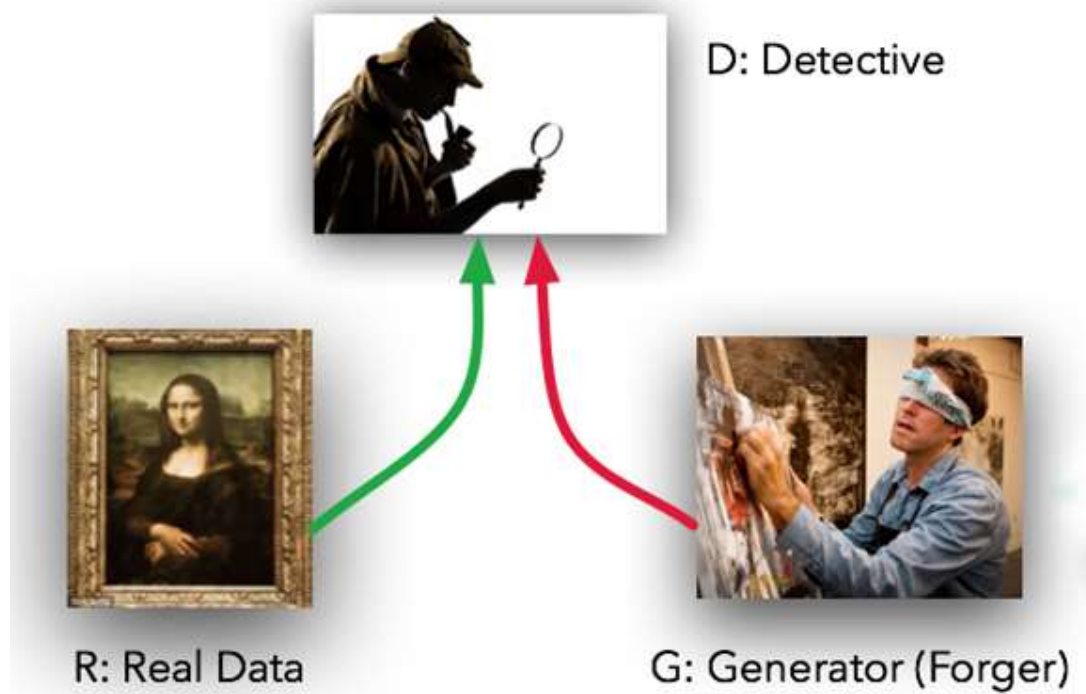
توابع فعال سازی غیر خطی دنبال می‌شود. این توابع مانند ReLU، Leaky ReLU یا tanh، غیرخطی بودن را وارد مدل می‌کنند. غیرخطی بودن برای یادگیری توزیع‌های پیچیده داده کاملاً ضروری است. بدون آن، شبکه فقط می‌تواند روابط خطی را بیاموزد و توانایی آن را برای تولید داده‌های واقعی و متنوع مانند تصاویر، که ذاتاً شامل روابط بسیار غیرخطی بین پیکسل‌ها می‌شود، به شدت محدود می‌کند.

از نظر ریاضی، عملکرد شبکه مولد را می‌توان به صورت $G(z; \theta_g)$ نشان داد. این نماد تأکید می‌کند که خروجی شبکه مولد تابعی از نویز ورودی z و پارامترهای شبکه مولد θ_g است. هدف نهایی شبکه مولد یادگیری مقادیر θ_g است به طوری که خروجی $G(z)$ از داده‌های واقعی قابل تشخیص نباشد. این بدان معنی است که توزیع داده‌های تولید شده باید با توزیع داده‌های واقعی مطابقت داشته باشد.

معماری خاص شبکه مولد به نوع داده تولید شده بستگی دارد. برای تولید تصویر، معمولاً از لایه‌های کانولوشنی استفاده می‌شود. این لایه‌ها به ویژه برای پردازش داده‌های تصویر مناسب هستند زیرا می‌توانند سلسله مراتب فضایی ویژگی‌ها را بیاموزند. برای مثال، لایه‌های کانولوشنی اولیه ممکن است ویژگی‌های ساده‌ای مانند لبه‌ها و گوشه‌ها را بیاموزند، در حالی که لایه‌های عمیق‌تر ویژگی‌های پیچیده‌تری مانند قطعات شی و بافت‌ها را می‌آموزند. این یادگیری سلسله مراتبی به مولد اجازه می‌دهد تا ساختار پیچیده تصاویر را ثبت کند و محتوای بصری واقعی و منسجمی تولید کند.

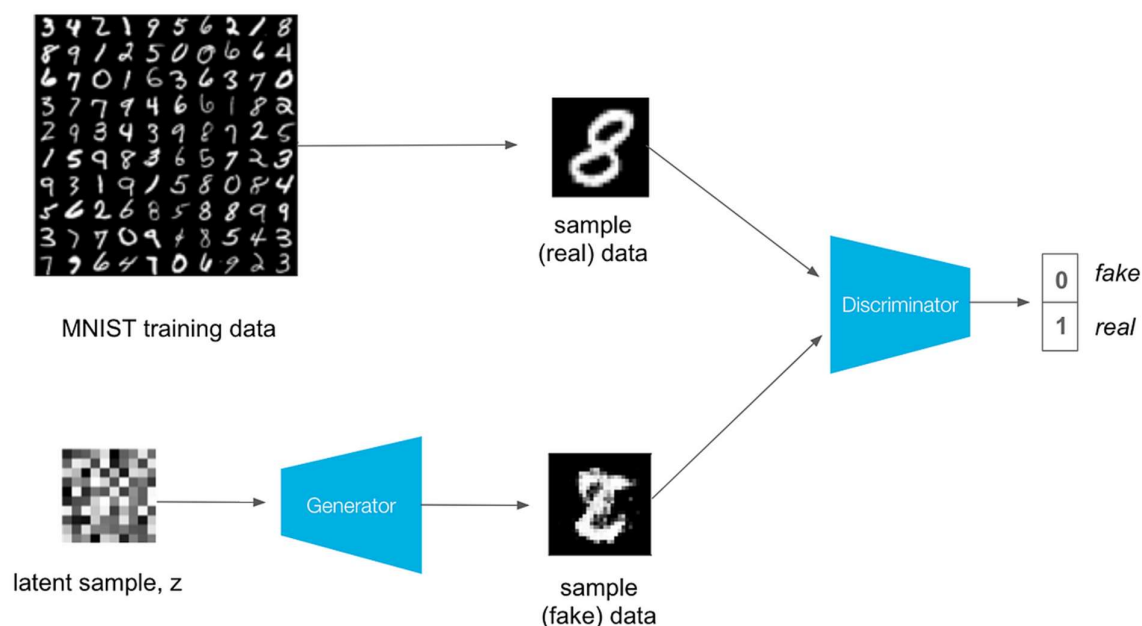
2-2 شبکه متمایزکننده

شبکه متمایزکننده که اغلب با D مشخص می‌شود، به عنوان منتقد یا داور هنر در چارچوب GAN عمل می‌کند. نقش اصلی آن تمایز بین نمونه‌های داده واقعی از مجموعه داده آموزشی و نمونه‌های داده جعلی تولید شده توسط مولد است. آن را به عنوان یک کارآگاه بسیار آموزش دیده در نظر بگیرید که سعی دارد ارزش تقلبی را شناسایی کند. متمایزکننده دو نوع ورودی دریافت می‌کند: داده واقعی که می‌توانیم به صورت x نمایش دهیم و داده جعلی $G(z)$ که توسط مولد تولید می‌شود.



شکل 2-2 تشخیص داده واقعی از ساختگی

خروجی متمایزکننده یک مقدار واحد است، معمولاً بین 0 و 1. این مقدار احتمال واقعی بودن نمونه ورودی را نشان می‌دهد. مقدار نزدیک به 1 نشان می‌دهد که متمایزکننده معتقد است ورودی واقعی است، در حالی که مقدار نزدیک به 0 نشان می‌دهد که فکر می‌کند ورودی جعلی است. هدف متمایزکننده تعیین دقیق این احتمالات است.



شکل 2-3 خروجی شبکه متمایزکننده

همانند مولد، متمایزکننده نیز یک شبکه عصبی است. معماری آن، مانند مولد، بسته به نوع داده‌ای که پردازش می‌شود، می‌تواند متفاوت باشد. برای داده‌های تصویر، معمولاً از لایه‌های کانولوشنی استفاده می‌شود. این لایه‌ها به متمایزکننده اجازه می‌دهد تا سلسله مراتب فضایی ویژگی‌ها را یاد بگیرد، درست مانند مولد، اما با هدفی متفاوت. در حالی که مولد از این ویژگی‌ها برای ایجاد تصاویر واقعی استفاده می‌کند، متمایزکننده از آن‌ها برای تجزیه و تحلیل تصاویر و شناسایی نشانه‌های ظریفی استفاده می‌کند که واقعی را از جعلی تشخیص می‌دهد.

متمایزکننده یاد می‌گیرد که ناسازگاری‌ها و مصنوعات را که ممکن است در تصاویر تولید شده وجود داشته باشد، تشخیص دهد. این موارد می‌تواند شامل ناهماهنگی در بافت‌ها، الگوهای غیرمعمول در اشکال اشیاء، توزیع رنگ غیرواقعی، یا سایر نقص‌های ظریف باشد که تصویر را مصنوعی بودن نشان می‌دهد. متمایزکننده اساساً در شناسایی "اثر انگشت" داده‌های تولید شده متخصص می‌شود.



شکل 2-4 ناسازگاری‌های به وجود آمده در زمان آموزش

از نظر ریاضی، خروجی متمایزکننده را می‌توان به صورت $D(x; \theta_d)$ برای داده‌های واقعی و $D(\theta_d; G(z))$ برای داده‌های جعلی نشان داد. در اینجا θ_d پارامترهای (وزن‌ها و بایاس‌ها) شبکه متمایزکننده را نشان می‌دهد. اینها پارامترهایی هستند که فرد متمایزکننده در طول آموزش یاد می‌گیرد تا به یک «منتقد» بهتر تبدیل شود.

هدف متمایزکننده به حداکثر رساندن توانایی خود در طبقه‌بندی صحیح نمونه‌های واقعی و جعلی است. در حالت ایده‌آل، باید احتمالات بالا را به داده‌های واقعی ($D(x)$ نزدیک به 1) و احتمالات پایین را به داده‌های جعلی ($D(G(z))$) نزدیک به 0 اختصاص دهد. عملکرد متمایزکننده برای موفقیت کلی GAN کاملاً حیاتی است. یک تمایز قوی یک سیگنال یادگیری قوی برای شبکه مولد فراهم می‌کند. اگر متمایزکننده در شناسایی تقلبی‌ها خوب باشد، مولد تحت فشار قرار می‌گیرد تا نمونه‌های واقعی‌تری ایجاد کند تا سعی کند آن را فریب دهد.

با این حال، یک تعادل ظریف وجود دارد. اگر متمایزکننده خیلی سریع خوب شود، می‌تواند یادگیری را برای مولد بسیار دشوار کند. این می‌تواند منجر به وضعیتی شود که در آن گرادینت‌های مولد بسیار کوچک می‌شوند (گرادینت‌های ناپدید شونده) که مانع از توانایی آن برای بهبود می‌شود. این یکی از چالش‌های آموزش GAN است و زمانی که به چالش‌های آموزشی بپردازیم، بیشتر درباره آن بحث خواهیم کرد. فشار و کشش مداوم بین مولد و متمایزکننده، اهداف رقابتی آن‌ها، موتوری است که فرآیند یادگیری GAN را به حرکت در می‌آورد. این یک رقابت پویا است که منجر به تولید داده‌های واقعی‌تر توسط مولد می‌شود و متمایزکننده به‌طور فزاینده‌ای قابل تشخیص می‌شود.

2-3 تابع ضرر

آموزش GAN‌ها یک عمل متعادل‌کننده ظریف است که توسط تعامل دو تابع ضرر رقیب هدایت می‌شود. این توابع ضرر به عنوان "قوانین بازی" عمل می‌کنند، عملکرد مولد و متمایزکننده را کمی می‌کنند و یادگیری آن‌ها را از طریق شیب نزول هدایت می‌کنند. آن‌ها تعریف می‌کنند که برای هر شبکه "برنده شدن" یا "باخت" در این بازی خصمانه به چه معناست.

مقاله اصلی GAN حداقل ضرر را معرفی کرد، مفهومی اساسی که ماهیت خصمانه آموزش GAN را تجسم می‌دهد. به صورت زیر بیان می‌شود:

$$\min_g \max_D V(D, G) = E_{\{x\}}[\log D(x, y)] + E_{\{z\}}[\log(1 - D(g(z, y), y))]$$

بیاید این معادله را تکه تکه باز کنیم.

- E_x نشان دهنده انتظار بیش از توزیع داده واقعی است. آن را به عنوان میانگین‌گیری نتایج در تمام نمونه‌های داده واقعی ممکن

- E_z نشان دهنده انتظار بیش از توزیع نویز است. این به معنای میانگین گیری از تمام بردارهای نویز تصادفی ممکن z .
- $D(x)$ خروجی متمایزکننده است که یک نمونه داده واقعی x داده می شود. در حالت ایده آل، متمایزکننده می خواهد این عدد نزدیک به 1 باشد، که نشان می دهد داده های واقعی را به درستی شناسایی می کند.
- $D(G(z))$ خروجی متمایزکننده است که یک نمونه داده جعلی $G(z)$ تولید شده توسط مولد داده می شود. متمایزکننده می خواهد این نزدیک به 0 باشد، که نشان می دهد به درستی داده های جعلی را شناسایی می کند.

حالا بیایید به دو عبارت موجود در حداقل ضرر نگاه کنیم:

- $\log D(x)$: این اصطلاح متمایزکننده را تشویق می کند تا داده های واقعی را به درستی طبقه بندی کند. با نزدیک شدن $D(x)$ به عدد 1، $\log D(x)$ افزایش می یابد. متمایزکننده می خواهد این اصطلاح را به حداکثر برساند.
- $\log(1 - D(G(z)))$: این اصطلاح تمایز کننده را تشویق می کند تا داده های جعلی را به درستی طبقه بندی کند. با نزدیک شدن $D(G(z))$ به 0، $\log(1 - D(G(z)))$ افزایش می یابد. متمایزکننده نیز می خواهد این اصطلاح را به حداکثر برساند.

با کنار هم قرار دادن اینها، هدف تمایز به حداکثر رساندن کل بیان است. می خواهد داده های واقعی و جعلی را به درستی شناسایی کند.

حال، بیایید دیدگاه مولد را در نظر بگیریم. می خواهد همین عبارت را به حداقل برساند. چرا؟ زیرا می خواهد $D(G(z))$ نزدیک به 1 باشد. اگر متمایزکننده فکر کند خروجی مولد واقعی است، مولد "برنده" است. توجه داشته باشید که مولد تنها بر ترم دوم، $\log(1 - D(G(z)))$ تاثیر دارد. برای به حداقل رساندن کل عبارت، مولد باید $D(G(z))$ را تا حد امکان بزرگ کند (نزدیک به 1).

حداقل ضرر نشان دهنده یک بازی با مجموع صفر است: آنچه یک شبکه به دست می آورد، دیگری از دست می دهد. این رابطه خصمانه قلب آموزش GAN است.

با این حال، حداقل ضرر یک نقطه ضعف دارد: ناپدید شدن گرادیان. به خصوص در مراحل اولیه آموزش، زمانی که مولد ضعیف است، متمایزکننده به راحتی می تواند نمونه های واقعی را از تقلبی تشخیص دهد. این بدان معناست که $D(G(z))$ نزدیک به 0 خواهد بود. مشکل این است که گرادیان $\log(1 - D(G(z)))$ نسبت به پارامترهای مولد زمانی که $D(G(z))$ بسیار کوچک می شود (نزدیک به 0 است. این امر یادگیری و بهبود را برای مولد دشوار می کند.

برای پرداختن به این موضوع، ضرر غیراشباع¹ برای مولد پیشنهاد شد:

$$\min_g E_{\{z\}}[\log D(G(z))]$$

به جای کمینه کردن $\log(1 - D(G(z)))$ ، مولد اکنون مستقیماً $\log D(G(z))$ را به حداکثر می‌رساند. این مولد را تشویق می‌کند تا نمونه‌هایی تولید کند که متمایزکننده به احتمال زیاد آن‌ها را به عنوان واقعی طبقه‌بندی کند. این ضرر غیر اشباع، گرادین‌های قوی‌تری را برای مولد فراهم می‌کند، به خصوص زمانی که عملکرد ضعیفی دارد، که در برخی موارد منجر به آموزش سریع‌تر و پایدارتر می‌شود.

در حالی که از دست دادن غیراشباع اغلب در عمل ترجیح داده شده است، راه حل کاملی نیست. گاهی اوقات می‌تواند منجر به فروپاشی حالت شود، وضعیتی که در آن مولد تعداد محدودی از خروجی‌ها را تولید می‌کند و تنها روی چند "حالت" توزیع داده تمرکز می‌کند.

به دلیل این چالش‌ها، محققان به طور مداوم در حال بررسی توابع ضرر جایگزین و استراتژی‌های آموزشی برای بهبود پایداری و عملکرد آموزش GAN هستند. جست‌وجو برای عملکرد خطا کامل برای GAN‌ها یک حوزه تحقیقاتی مداوم است.

2-4 فرآیند آموزش

آموزش GAN یک عمل متعادل‌کننده² ظریف است، یک جنگ طناب‌کشی بین مولد و متمایزکننده. این یک فرآیند تکراری است که در آن هر دو شبکه به طور همزمان آموزش داده می‌شوند، اما با اهداف متضاد. هدف نهایی رسیدن به تعادل نش³ است. از نظر تئوری بازی، این حالت پایداری است که در آن نه مولد و نه متمایزکننده نمی‌توانند عملکرد خود را با تغییر یک طرفه استراتژی خود (یعنی تنظیم پارامترهای آن) بهبود بخشند در حالی که استراتژی شبکه دیگر بدون تغییر باقی می‌ماند. به عبارت ساده‌تر، این نقطه‌ای است که مولد بهترین داده‌های جعلی ممکن را تولید می‌کند، و متمایزکننده بهترین کار ممکن را برای تشخیص واقعی از جعلی انجام می‌دهد، و هیچ‌کدام نمی‌توانند به تنهایی بهتر شوند.

با این حال، دستیابی به یک تعادل کامل نش در عمل بسیار چالش برانگیز است. چشم انداز اهداف رقابتی پیچیده است و فرآیند آموزش اغلب شامل پیمایش یک مسیر ظریف برای اجتناب از دام‌های مختلف است.

فرآیند آموزش معمولاً در مینی دسته‌ها انجام می‌شود. به جای پردازش کل مجموعه داده‌ها به یکباره، که از نظر محاسباتی بسیار گران است، داده‌ها به دسته‌های کوچک‌تر تقسیم می‌شوند. در هر تکرار آموزشی، یک دسته

¹non-saturating

²ballancing

³Nash equilibrium

کوچک از نمونه‌های داده واقعی به طور تصادفی از مجموعه داده آموزشی استخراج می‌شود و یک دسته کوچک از بردارهای نویز تصادفی تولید می‌شود. این مینی دسته‌ها برای محاسبه خطا و به روزرسانی پارامترهای شبکه استفاده می‌شوند.

فرآیند آموزش به طور متناوب بین به‌روزرسانی متمایزکننده و مولد انجام می‌شود. ابتدا، متمایزکننده به روز می‌شود. هم مینی دسته‌ها از نمونه‌های داده واقعی و هم مینی دسته از نمونه‌های جعلی تولید شده توسط متمایزکننده را دریافت می‌کند. متمایزکننده میزان خطای ضرر خود را بر اساس میزان توانایی تشخیص داده‌های واقعی و جعلی محاسبه می‌کند. همانطور که قبلاً بحث شد، از حداقل خطا یا ضرر غیراشباع استفاده می‌کند. سپس، پارامترهای خود (وزن‌ها و بایاس‌ها) را با استفاده از گرادیان نزول به‌روزرسانی می‌کند. گرادیان نزول یک الگوریتم بهینه‌سازی است که به طور مکرر پارامترها را در جهتی تنظیم می‌کند که تابع ضرر را به حداقل می‌رساند. هدف به‌روزرسانی متمایزکننده به حداقل رساندن ضرر آن است، به این معنی که می‌خواهد در طبقه‌بندی صحیح نمونه‌های واقعی و جعلی بهتر شود. بعد، مولد به روز می‌شود. مهم‌تر از همه، مولد یک مینی دسته جدید از بردارهای نویز تصادفی دریافت می‌کند. این نمونه‌های جعلی مربوطه را با استفاده از پارامترهای فعلی خود تولید می‌کند. سپس مولد خطای خود را با استفاده از حداقل خطا یا خطای غیراشباع محاسبه می‌کند. هدف مولد این است که خطای خود را به حداقل برساند (یا در مورد خطای غیر اشباع، احتمال گول خوردن متمایزکننده را به حداکثر برساند). برای رسیدن به این هدف، پارامترهای خود را با استفاده از گرادیان نزول به روز می‌کند. یک نکته بسیار مهم این است که در طول به روز رسانی مولد، پارامترهای متمایزکننده ثابت نگه داشته می‌شوند. این امر ضروری است زیرا تضمین می‌کند که مولد در حال یادگیری بهبود خود نسبت به وضعیت فعلی متمایزکننده است. اگر پارامترهای متمایزکننده نیز در طول به‌روزرسانی مولد تغییر می‌کردند، مولد یک هدف متحرک را تعقیب می‌کرد و فرآیند یادگیری را ناپایدار می‌کرد.

این فرآیند متناوب به‌روزرسانی متمایزکننده و مولد برای تعداد معینی از دوره‌ها تکرار می‌شود. آموزش تا رسیدن به سطح مطلوبی از عملکرد (به عنوان مثال، نمونه‌های تولید شده به اندازه کافی واقعی به نظر می‌رسند) یا تا زمانی که منابع محاسباتی تمام شوند ادامه می‌یابد.

در طول آموزش، نظارت بر عملکرد هر دو شبکه کاملاً حیاتی است. اگر متمایزکننده خیلی سریع قوی شود، مولد ممکن است برای یادگیری مشکل داشته باشد که منجر به از بین رفتن گرادیان‌ها می‌شود. این به این دلیل است که متمایزکننده در شناسایی تقلبی‌ها آنقدر خوب می‌شود که مولد بازخورد مفید بسیار کمی برای بهبود دریافت می‌کند. برعکس، اگر مولد خیلی سریع قوی شود، می‌تواند منجر به فروپاشی حالت شود. در این شرایط، مولد در تولید انواع محدودی از خروجی‌ها گیر می‌کند، حتی اگر داده‌های واقعی تنوع بسیار بیشتری دارند. اساساً چند «ترفند» برای فریب دادن متمایزکننده پیدا می‌کند و روی آن‌ها متمرکز می‌شود و بخش‌های دیگر توزیع داده‌ها را نادیده می‌گیرد.

برای کاهش این چالش‌ها و بهبود فرآیند آموزش، تکنیک‌های مختلفی به کار گرفته می‌شود. این موارد عبارتند از:

تنظیم نرخ یادگیری: نرخ یادگیری سرعت به‌روزرسانی پارامترهای شبکه‌ها را کنترل می‌کند. تنظیم دقیق نرخ یادگیری برای آموزش پایدار ضروری است.

برای کاهش این چالش‌ها و بهبود فرآیند آموزش، تکنیک‌های مختلفی به کار گرفته می‌شود. این موارد عبارتند از:

- **تنظیم نرخ یادگیری:** نرخ یادگیری سرعت به‌روزرسانی پارامترهای شبکه‌ها را کنترل می‌کند. تنظیم دقیق نرخ یادگیری برای آموزش پایدار ضروری است.
- **استفاده از بهینه‌سازهای مختلف:** الگوریتم‌های بهینه‌سازی مختلف (مانند Adam, RMSprop) می‌توانند ویژگی‌های همگرایی متفاوتی داشته باشند و بر پویایی آموزش تأثیر بگذارند.
- **استفاده از روش‌های منظم‌سازی¹:** تکنیک‌های منظم‌سازی می‌تواند به جلوگیری از برازش بیش از حد و بهبود تعمیم شبکه‌ها کمک کند.

آموزش GAN‌ها اغلب بیشتر به عنوان یک هنر توصیف می‌شود تا یک علم. برای دستیابی به نتایج خوب نیاز به تنظیم، نظارت و آزمایش دقیق دارد. هیچ "دستور العمل جادویی" واحدی وجود ندارد که برای همه سناریوهای آموزشی GAN کار کند. محققان به طور مداوم در حال بررسی تکنیک‌های آموزشی جدید و بهبود یافته برای پایداری و قابل اطمینان‌تر کردن آموزش GAN هستند.

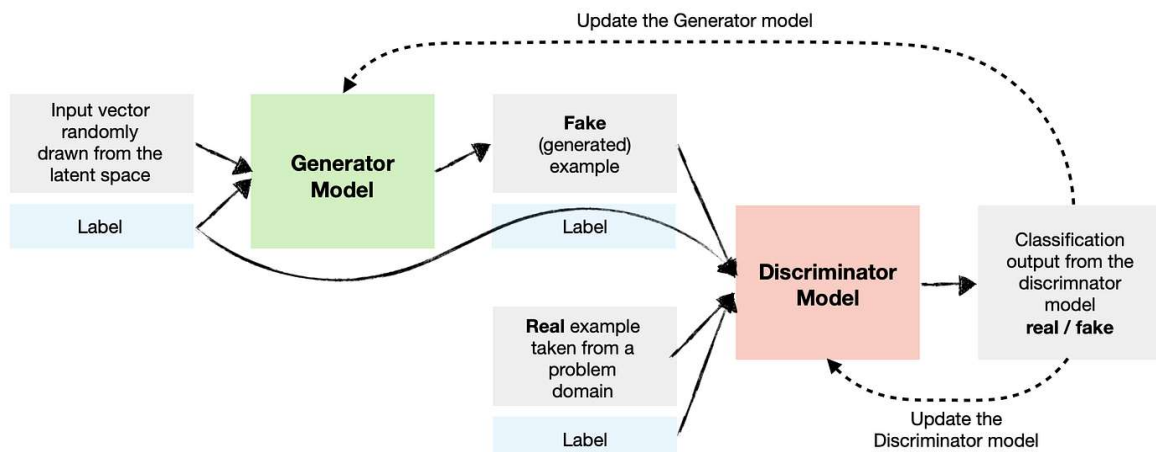
¹Regularization methods

فصل 3: انواع GAN

1-3 شبکه گن شرطی

همانطور که در مورد GAN های سنتی بحث کردیم، یاد می گیرند که داده ها را از نویز تصادفی تولید کنند. اگرچه این رویکرد قابل توجه است، اما کنترل محدودی بر فرآیند تولید ارائه می دهد. ما نمی توانیم مستقیماً مشخص کنیم که می خواهیم GAN چه نوع داده ای را تولید کند. برای مثال، اگر یک GAN را روی تصاویر چهره ها آموزش دهیم، چهره های مختلفی تولید می کند، اما نمی توانیم بگوییم که چهره ای با ویژگی های خاصی مانند چشم های آبی، لبخند یا عینک تولید کند. اینجاست که GAN های شرطی (CGAN) وارد می شوند.

گن های شرطی (CGAN)¹ [2] چارچوب اصلی GAN را با ترکیب اطلاعات مشروط در مولد و متمایز کننده گسترش می دهند. این اطلاعات شرطی که اغلب با y مشخص می شود، به ما اجازه می دهد تا فرآیند تولید را هدایت کنیم و داده هایی با ویژگی های خاص تولید کنیم. در نظر بگیرید که به هنرمند (مولد) مجموعه ای از دستورالعمل ها یا توضیحی در مورد اینکه چه چیزی نقاشی کند، و به منتقد هنری (تمایز کننده) اطلاعاتی مشابه بدهید تا بتوانند نقاشی را بر اساس آن قضاوت کنند.



شکل 3-1 نوع ساختار conditional gan

اطلاعات شرطی y می تواند هر چیزی مربوط به داده های تولید شده باشد. برای تولید تصویر، می تواند برچسب های کلاس (مانند «گربه»، «سگ»، «پرنده»، توضیحات متنی («ماشین قرمز پارک شده در کنار ساحل»)، یا حتی تصاویر دیگر (برای ترجمه تصویر به تصویر) باشد. برای انواع دیگر داده ها، این می تواند انواع مختلف برچسب یا ابرداده² باشد.

¹ Conditional GANs

² Metadata

نحوه ادغام این اطلاعات مشروط بسیار مهم است. در یک CGAN، هم مولد و هم متمایزکننده y را علاوه بر بردار نویز معمول z (برای مولد) و نمونه داده واقعی یا تولید شده (برای متمایزکننده) به عنوان ورودی دریافت می‌کنند.

• **مولد:** ورودی مولد ترکیبی از بردار نویز z و اطلاعات شرطی y می‌شود. این شبکه یاد می‌گیرد که داده‌های مشروط بر y تولید کند. از نظر ریاضی، می‌توانیم خروجی مولد را به صورت $G(y, z; \theta_g)$ نشان دهیم. مولد سعی می‌کند داده‌هایی تولید کند که نه تنها واقعی به نظر می‌رسند، بلکه با شرط y نیز مطابقت دارند.

• **تمتایزکننده:** متمایزکننده همچنین اطلاعات شرطی y را به همراه نمونه داده واقعی یا تولید شده به عنوان ورودی دریافت می‌کند. وظیفه آن نه تنها تمایز بین داده‌های واقعی و جعلی است، بلکه اطمینان از مطابقت داده‌های تولید شده با شرایط داده شده است. از نظر ریاضی، خروجی متمایزکننده را می‌توان به صورت $D(y, x; \theta_d)$ برای داده‌های واقعی و $G(z)D(y, y; \theta_d)$ برای داده‌های جعلی نشان داد. متمایزکننده یاد می‌گیرد که قضاوت کند آیا داده‌ها واقعی هستند و آیا با شرایط مطابقت دارند یا خیر.

توابع خطا برای CGANها برای ترکیب اطلاعات شرطی اصلاح شده‌اند. به عنوان مثال، حداقل خطا به صورت زیر می‌شود:

$$\min_g \max_D V(D, G) = E_{\{x, y\}} [\log D(x, y)] + E_{\{z, y\}} [\log (1 - D(G(z, y), y))]$$

توجه داشته باشید که y در هر دو عبارت ظاهر می‌شود، که نشان می‌دهد هم مولد و هم متمایزکننده مشروط به این اطلاعات هستند. خطای غیر اشباع نیز می‌تواند به طور مشابه اصلاح شود.

تفاوت فرآیند آموزش برای CGANها با GANهای معمولی، مرحله اضافی ارائه اطلاعات شرطی y به هر دو شبکه در طول هر تکرار است.

CGANها با فعال کردن تولید داده‌های کنترل شده، طیف گسترده‌ای از برنامه‌ها را باز کرده‌اند. برخی از نمونه‌های برجسته عبارتند از:

• **ترکیب تصویر با برچسب‌های کلاس:** ایجاد تصاویری از اشیاء یا صحنه‌های خاص با ارائه برچسب‌های مربوطه.

• **ویرایش و دستکاری تصویر:** اصلاح تصاویر موجود بر اساس اطلاعات شرطی، مانند تغییر رنگ یک شی یا افزودن عناصر جدید.

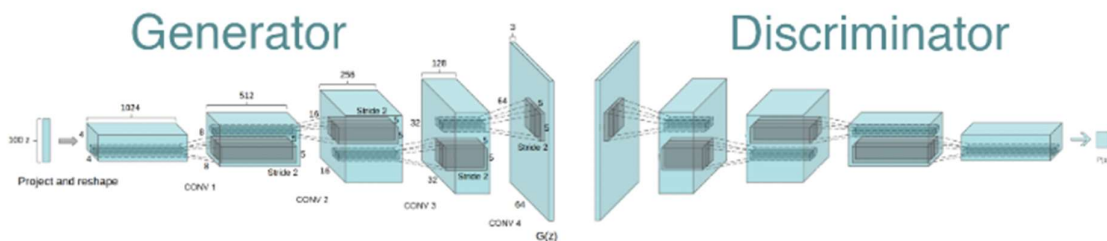
• **ترجمه تصویر به تصویر:** تبدیل تصاویر از یک دامنه به دامنه دیگر بر اساس داده‌های آموزشی جفت یا جفت نشده. به عنوان مثال، تبدیل طرح به عکس یا روز به شب.

• سنتز متن به تصویر: ایجاد تصاویر از توضیحات متن.

2-3 GANهای کانولوشنال عمیق

GAN های کانولوشنال عمیق¹ (DCGANs) [3] یک پیشرفت قابل توجه در تاریخ GAN ها، به ویژه در حوزه تولید تصویر است. در حالی که مقاله اصلی GAN پتانسیل آموزش خصمانه را نشان داد، کیفیت و وضوح تصاویر تولید شده اغلب محدود بود. DCGAN ها این محدودیت ها را با گنجانیدن قدرت شبکه های عصبی کانولوشنی (CNN) در معماری GAN برطرف کردند. این تغییر به ظاهر ساده تأثیر عمیقی داشت و منجر به بهبود چشمگیر کیفیت تصویر شد و راه را برای بسیاری از پیشرفت های بعدی در تحقیقات GAN هموار کرد.

نوآوری کلیدی DCGAN ها استفاده سیستماتیک از لایه های کانولوشنی در هر دو مولد و متمایز کننده است. CNN ها به ویژه برای وظایف پردازش تصویر مناسب هستند زیرا می توانند سلسله مراتب فضایی ویژگی ها را بیاموزند. لایه های کانولوشنی یاد می گیرند که الگوها را در مقیاس های مختلف تشخیص دهند، از لبه ها و بافت های ساده در لایه های اولیه گرفته تا اجزای شی پیچیده و ترکیب بندی ها در لایه های عمیق تر. این یادگیری ویژگی سلسله مراتبی برای ثبت ساختار پیچیده تصاویر و تولید محتوای بصری واقعی بسیار مهم است.



شکل 2-3 ساختار شبکه dcgan

DCGAN ها مجموعه ای از دستورالعمل های معماری را برای ساختن GAN های کانولوشنی پایدار و موثر معرفی کردند. این دستورالعمل ها شامل:

- جایگزینی لایه های ادغام² با پیچش های گام به گام³ در متمایز کننده: CNN های سنتی اغلب از لایه های ادغام (مانند max pooling) برای پایین آوردن نقشه های ویژگی استفاده می کنند. DCGAN ها دریافتند که پیچش های گام به گام، جایی که عملیات پیچش خود نمونه برداری را انجام می دهد، در متمایز کننده بهتر عمل می کند.

¹ Deep Convolutional GANs

² Pooling layers

³ Stride Convolutions

• استفاده از کانولوشن‌های جابجا شده^۱ (دکانولوشن) در مولد: مولد برای تولید یک تصویر باید بردار نویز را نمونه‌برداری کند. DCGAN‌ها برای این منظور از کانولوشن‌های جابجا شده، که به عنوان دکانولوشن^۲ نیز شناخته می‌شوند، استفاده می‌کنند. این لایه‌ها عملاً عملیات پیچیدگی را «معکوس» می‌کنند و به مولد اجازه می‌دهند که به تدریج ابعاد فضایی نقشه‌های ویژگی را افزایش دهد و تصاویری با وضوح بالاتر تولید کند.

• نرمال‌سازی دسته‌ای^۳: نرمال‌سازی دسته‌ای تکنیکی است که فعال‌سازی هر لایه را عادی می‌کند، که به تثبیت آموزش کمک می‌کند و امکان نرخ یادگیری بالاتر را فراهم می‌کند. DCGAN‌ها نرمال‌سازی دسته‌ای را هم در مولد و هم در متمایزکننده گنجانده‌اند.

• فعال‌ساز ReLU در مولد (به جز لایه خروجی): ReLU و انواع آن توابع فعال‌سازی محبوب در یادگیری عمیق هستند. DCGAN‌ها از ReLU در تمام لایه‌های مولد استفاده می‌کردند به جز لایه خروجی، جایی که از tanh استفاده می‌کردند.

• فعال‌سازی Leaky ReLU در متمایزکننده: Leaky ReLU گونه‌ای از ReLU است که به مشکل "ReLU در حال مرگ"^۴ می‌پردازد. DCGAN‌ها از Leaky ReLU در متمایزکننده استفاده کردند.

• حذف لایه‌های کاملاً متصل: DCGAN‌ها معماری‌های کاملاً کانولوشنال را ترجیح می‌دهند و در صورت امکان لایه‌های کاملاً متصل را حذف می‌کنند. این امکان پردازش مستقیم و کارآمدتر داده‌های تصویر را فراهم کرد.

با رعایت این دستورالعمل‌های معماری، DCGAN‌ها توانایی قابل توجهی در تولید تصاویر با کیفیت بالا نشان دادند. آن‌ها نشان دادند که GAN‌های کانولوشنی می‌توانند یاد بگیرند که توزیع‌های پیچیده تصویر را ضبط کنند و محتوای بصری واقعی و منسجم تولید کنند.

تاثیر DCGAN بسیار زیاد بود. آن‌ها نه تنها کیفیت تصاویر تولید شده توسط GAN را به طور قابل توجهی بهبود بخشیدند، بلکه چارچوب پایدارتر و قابل اعتمادتری را برای آموزش GAN فراهم کردند. DCGAN‌ها به یک معماری استاندارد برای وظایف تولید تصویر تبدیل شدند و به عنوان پایه‌ای برای بسیاری از نوآوری‌های بعدی GAN عمل کردند. آن‌ها راه را برای تحقیق در مورد معماری‌های GAN پیشرفته‌تر، مانند Progressive GAN و StyleGAN که در بخش‌های بعدی توضیح داده خواهم شد، هموار کردند که مرزهای سنتز تصویر را بیشتر می‌کرد.

^۱ Transposed Convolutions
^۲ deconvolutions
^۳ Batch Normalization
^۴ Dying ReLU

3-3 شبکه [4] Wasserstein GANs (WGANs)

Wasserstein GANs (WGANs) یک پیشرفت مهم در رسیدگی به بی‌ثباتی آموزشی است که معماری‌های GAN قبلی را آزار می‌داد. GAN‌های سنتی، در حالی که قادر به تولید نتایج چشمگیر هستند، اغلب از مشکلاتی مانند ناپدید شدن گرادین‌ها، فروپاشی حالت، و فقدان معیارهای همگرایی قابل اعتماد رنج می‌برند. WGAN‌ها با معرفی یک تابع خطای جدید بر اساس فاصله Wasserstein که به عنوان فاصله حرکت دهنده زمین¹ نیز شناخته می‌شود، با این چالش‌ها مقابله کردند.

ایده اصلی پشت WGAN‌ها جایگزینی توابع خطای سنتی (حداقل یا غیراشباع) با یک تابع ضرر است که مستقیماً فاصله Wasserstein بین توزیع داده واقعی و توزیع داده تولید شده را تقریب می‌کند. فاصله Wasserstein اندازه‌گیری معنادارتری از فاصله بین دو توزیع احتمال را در مقایسه با واگرایی‌های مورد استفاده در خطای GAN سنتی ارائه می‌دهد. اساساً حداقل "تلاش" مورد نیاز برای تبدیل یک توزیع به توزیع دیگر، مانند حرکت زمین برای پر کردن حفره‌ها را کمیت می‌کند.

چرا این مهم است؟ فاصله Wasserstein چندین ویژگی مطلوب دارد که آن را برای آموزش GAN مناسب می‌کند:

- **گرادین‌های با شیب ملایم‌تر:** برخلاف خطای GAN سنتی، فاصله Wasserstein حتی زمانی که توزیع‌ها از هم دور هستند، گرادین‌های صاف را فراهم می‌کند. این به کاهش مشکل گرادین ناپدید شدن کمک می‌کند و به مولد اجازه می‌دهد حتی در مراحل اولیه آموزش به طور موثر یاد بگیرد.

- **فروپاشی حالت کاهش یافته:** فاصله Wasserstein در مقایسه با خطای GAN سنتی کمتر مستعد فروپاشی حالت است. این مولد را تشویق می‌کند تا تنوع کامل توزیع داده‌ها را به جای گیرکردن در چند حالت، ثبت کند.

- **متریک همگرایی معنادار:** فاصله Wasserstein یک معیار معنادار برای ارزیابی پیشرفت آموزش GAN ارائه می‌دهد. ما می‌توانیم فاصله Wasserstein را در طول آموزش کنترل کنیم تا میزان یادگیری مولد را ردیابی کنیم و تعیین کنیم که چه زمانی آموزش همگرا شده است.

با این حال، محاسبه فاصله Wasserstein دقیق از نظر محاسباتی چالش برانگیز است. WGAN‌ها از دوگانگی Kantorovich-Rubinstein برای تقریب فاصله Wasserstein با استفاده از یک فرمول دوگانه استفاده می‌کنند. این فرمول‌بندی دوگانه شامل آموزش یک شبکه «منتقد» است، که نقشی مشابه با متمایزکننده در GAN‌های سنتی

¹Earth Mover Distance

بازی می‌کند، اما با یک تفاوت اساسی: منتقد برای طبقه‌بندی واقعی در مقابل جعلی آموزش دیده نیست. در عوض، می‌آموزد که تابعی را تخمین بزند که گرادیان آن کران پایین‌تری را در فاصله Wasserstein فراهم می‌کند.

تابع ضرر برای WGAN‌ها که از دوگانگی کانتورویچ-روبینشتاین مشتق شده است، به صورت زیر است:

$$L_{wgan} = E_{\{x\}}[f(x)] - E_{\{z\}}[f(G(x))]$$

که در آن f تابعی است که شبکه انتقادی یاد می‌گیرد. منتقد سعی می‌کند این ضرر را به حداکثر برساند، در حالی که مولد تلاش می‌کند آن را به حداقل برساند.

یک چالش کلیدی در آموزش WGAN‌ها، اعمال یک محدودیت Lipschitz در تابع انتقادی است. این محدودیت برای اطمینان از اینکه فرمول دوتایی دقیقاً فاصله Wasserstein را تقریب می‌کند ضروری است. مقاله اصلی WGAN برش وزن را به عنوان روشی برای اعمال محدودیت Lipschitz پیشنهاد کرد، که در آن وزن‌های شبکه انتقادی پس از هر به‌روزرسانی به محدوده خاصی بریده می‌شوند.

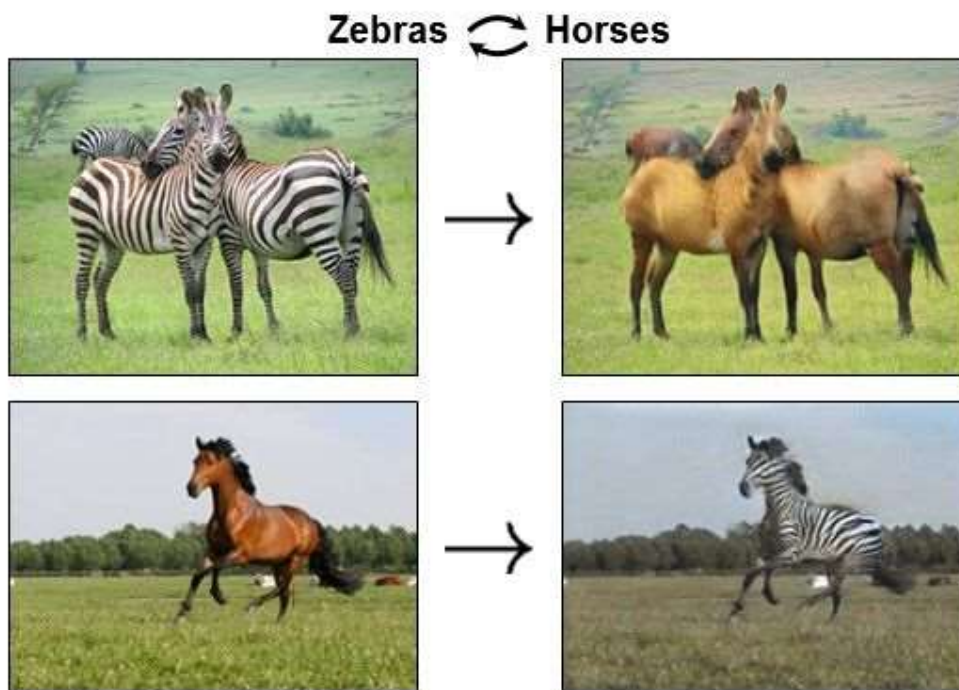
با این حال، برش وزن محدودیت‌های خود را دارد. اگر محدوده برش خیلی کوچک باشد، می‌تواند منجر به ناپدید شدن گرادیان‌ها شود و اگر محدوده برش خیلی زیاد باشد، می‌تواند باعث ناپایداری شود. نسخه‌های بهبود یافته WGAN‌ها، مانند WGAN-GP (Wasserstein GAN با Gradient Penalty)، روش‌های جایگزینی را برای اعمال محدودیت Lipschitz، مانند افزودن یک جریمه گرادیان به تابع ضرر، پیشنهاد کرده‌اند.

WGAN‌ها به طور قابل توجهی ثبات و قابلیت اطمینان آموزش GAN را بهبود بخشیده‌اند. آن‌ها آموزش GAN‌ها را بر روی مجموعه داده‌های پیچیده‌تر و تولید نمونه‌های با کیفیت بالاتر و متنوع‌تر ممکن کرده‌اند. در حالی که WGAN‌ها مجموعه‌ای از چالش‌های خاص خود را دارند، آن‌ها یک گام بزرگ به جلو در تلاش برای تکنیک‌های آموزشی قوی‌تر و موثرتر GAN هستند.

3-4 شبکه‌های متخاصم مولد چرخه

شبکه‌های متخاصم مولد چرخه¹ یا CycleGAN‌ها [5] یک راه حل قدرتمند و ظریف برای مشکل ترجمه تصویر به تصویر بدون نیاز به داده‌های آموزشی جفتی ارائه می‌دهند. روش‌های سنتی ترجمه تصویر به تصویر اغلب بر مجموعه داده‌های جفتی تکیه می‌کنند، که در آن هر تصویر در دامنه منبع با یک تصویر مربوطه در دامنه هدف مطابقت دارد (به عنوان مثال، عکس یک اسب با عکس یک گورخر). با این حال، به دست آوردن چنین مجموعه داده‌های جفتی می‌تواند در بسیاری از سناریوهای دنیای واقعی گران، زمان بر یا حتی غیرممکن باشد.

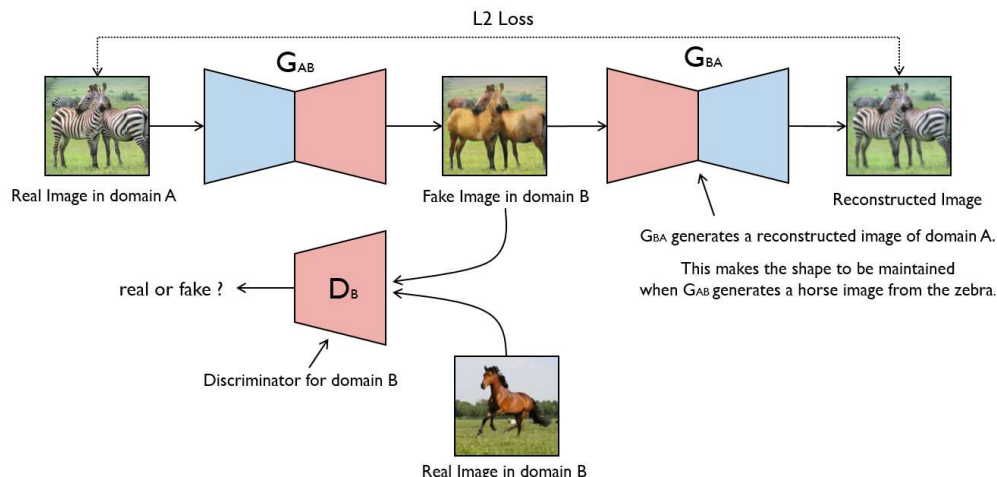
¹Cycle Generative Adversarial Networks



شکل 3-3 نمونه تصویر ساخته شده توسط *cyclegan*

CycleGAN ها با یادگیری ترجمه بین دو دامنه فقط با استفاده از داده‌های جفت نشده، هوشمندانه این نیاز را دور می‌زنند. آن‌ها از قدرت آموزش خصمانه استفاده می‌کنند و از خطای ثبات چرخه را برای اطمینان از این که نگاشت‌های آموخته شده سازگار و معنادار هستند، معرفی می‌کنند.

تصور کنید می‌خواهید ترجمه اسب‌ها را به گورخر و بالعکس یاد بگیرید، اما فقط مجموعه‌ای از عکس‌های اسب و مجموعه‌ای جداگانه از عکس‌های گورخر دارید. شما هیچ عکسی از اسب‌هایی که به طور خاص به گورخر تبدیل شده اند ندارید. CycleGAN ها می‌توانند این ترجمه را با اعمال این محدودیت یاد بگیرند که اگر اسبی را به گورخر ترجمه کنید و سپس گورخر را دوباره به اسب ترجمه کنید، باید چیزی شبیه به اسب اصلی دریافت کنید. این جوهر ثبات چرخه است.



شکل 3-4 ساختار شبکه cyclegan

CycleGAN از دو مولد و دو متمایزکننده تشکیل شده است:

- **مولد Gab:** تصاویر را از دامنه X (به عنوان مثال اسب) به دامنه Y (مثلاً گورخرها) برمی گرداند.
 - **مولد Gba:** تصاویر را از دامنه Y (مثلاً گورخرها) به دامنه X (مثلاً اسب) برمی گرداند.
 - **متمایزکننده Db:** بین تصاویر واقعی در دامنه Y و تصاویر جعلی تولید شده توسط G تمایز قائل می شود.
 - **متمایزکننده Da:** بین تصاویر واقعی در دامنه X و تصاویر جعلی تولید شده توسط F تمایز قائل می شود.
- فرآیند آموزش شامل دو ضرر خصمانه و یک ضرر ثبات چرخه است:

1. **ضرر خصمانه برای Gab:** این ضرر G را تشویق می کند تا تصاویری شبیه گورخرهای واقعی تولید کند و D_y را فریب دهد.

2. **ضرر خصمانه برای Gba:** این ضرر F را تشویق می کند تا تصاویری شبیه اسبهای واقعی تولید کند و D_x را فریب دهد.

3. **ضرر ثبات چرخه:** این افت ترکیب دو مولد $(F(G(x)))$ را تشویق می کند که به تصویر اصلی x نزدیک شود و به طور مشابه، $G(F(y))$ نزدیک به y باشد. این تضمین می کند که ترجمه ها سازگار هستند و مولدها نداشت های معنی دار بین دو حوزه را یاد می گیرند.

ضرر ثبات چرخه برای یادگیری بدون داده های جفتی بسیار مهم است. این به عنوان نوعی نظارت بر خود عمل می کند و مولدها را برای یادگیری تحولاتی هدایت می کند که برگشت پذیر و سازگار هستند. دو نوع ضرر ثبات چرخه وجود دارد:

• ثبات چرخه رو به جلو: $|F(G(x)) - x|$ باید کوچک باشد

• ثبات چرخه به عقب: $|G(F(y)) - y|$ باید کوچک باشد

ضرر کل برای CycleGAN ترکیبی از دو ضرر خصمانه و ضرر ثبات چرخه است. فرآیند آموزش شامل آموزش همزمان مولدها و تمایزکنندگان برای به حداقل رساندن این ضرر ترکیبی است.

CycleGANها با موفقیت در طیف گسترده‌ای از وظایف ترجمه تصویر به تصویر اعمال شده‌اند، از جمله:

• **تغییر شکل شی^۱:** تبدیل اسب‌ها به گورخر، گربه‌ها به سگ و غیره.

• **انتقال سبک^۲:** انتقال سبک هنری یک تصویر به تصویر دیگر.

• **بهبود عکس:** بهبود کیفیت یا ظاهر عکس‌ها.

• **تطبیق دامنه:** تطبیق مدل‌های آموزش دیده در یک دامنه به دامنه دیگر.

CycleGANها قدرت آموزش خصمانه و ثبات چرخه را برای یادگیری نگاشت‌های پیچیده بین دامنه‌ها بدون تکیه بر داده‌های آموزشی جفتی نشان داده‌اند. آن‌ها فرصت‌های جدیدی را برای ترجمه تصویر به تصویر باز کرده‌اند و الهام‌بخش تحقیقات بیشتر در مورد روش‌های یادگیری بدون نظارت و با نظارت ضعیف هستند.

5-3 شبکه StyleGAN

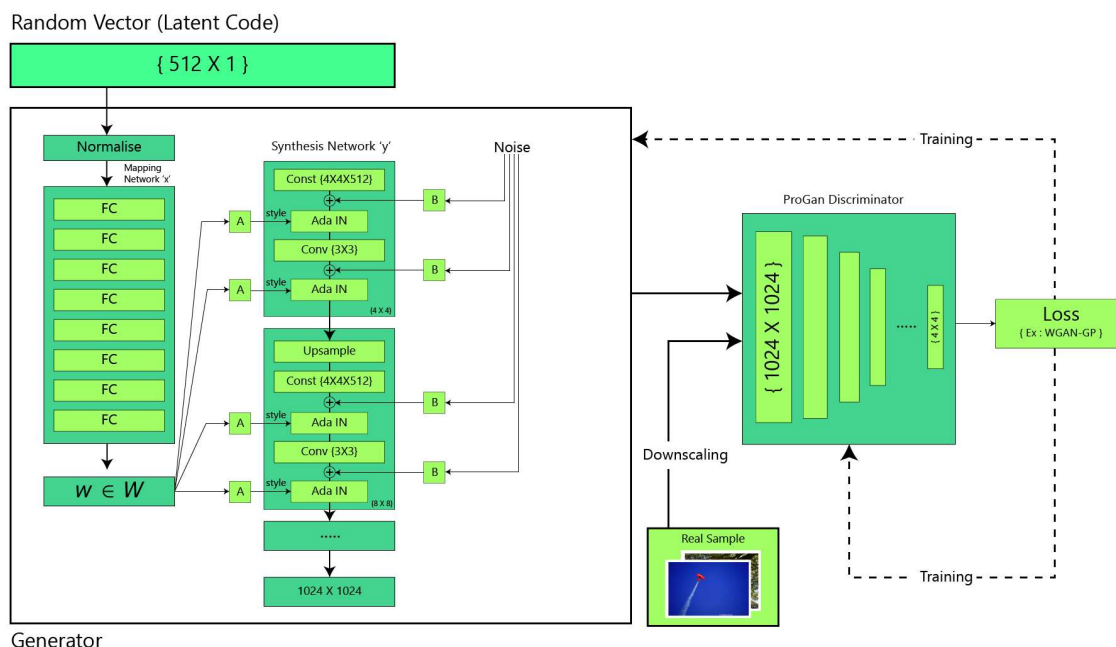
شبکه StyleGAN[6]^۳، توسعه یافته توسط انویدیا، نشان‌دهنده یک جهش بزرگ به جلو در زمینه ترکیب تصاویر، به ویژه برای تولید تصاویر با وضوح بالا و واقعی از چهره‌ها است. StyleGANها نه تنها تصاویر بسیار واقعی را تولید می‌کنند، بلکه کنترل بی‌سابقه‌ای بر محتوای تولید شده نیز ارائه می‌دهند که امکان دستکاری دقیق ویژگی‌های چهره، حالات و سبک‌ها را فراهم می‌کند.

^۱ Object Transfiguration
^۲ Style Transfer
^۳



شکل 3-5 چند نمونه تصویر ساخته شده توسط stylegan

نوآوری کلیدی StyleGAN ها در معماری مولد جدید آن نهفته است که مفهوم مدولاسیون سبک را معرفی می‌کند. StyleGAN به جای تغذیه مستقیم بردار نویز به مولد، ابتدا بردار نویز (و به صورت اختیاری، یک بردار شرطی‌سازی) را با استفاده از یک شبکه نگاشت آموخته شده در یک بردار سبک متوسط ترسیم می‌کند. سپس این بردار سبک فرآیند تولید تصویر را در سطوح مختلف جزئیات کنترل می‌کند.



شکل 3-6 ساختار شبکه stylegan

مولد در StyleGAN ها به عنوان یک شبکه در حال رشد پیشرو طراحی شده است. با تولید یک تصویر با وضوح بسیار پایین شروع می‌شود و به تدریج لایه‌هایی را برای افزایش وضوح اضافه می‌کند و جزئیات را در هر مرحله

اصلاح می‌کند. بردار سبک در چندین نقطه به مولد، به طور خاص پس از هر لایه کانولوشن، تزریق می‌شود. این کار با مدل کردن هسته‌های کانولوشنی با استفاده از بردار سبک^۱ انجام می‌شود. بردار سبک به طور موثر قدرت عملیات کانولوشن را کنترل می‌کند و بر ویژگی‌های آموخته شده توسط هر لایه تأثیر می‌گذارد.

بخش‌های مختلف بردار سبک، جنبه‌های مختلف تصویر تولید شده را کنترل می‌کنند. برخی از قسمت‌ها ممکن است ویژگی‌های درشتی مانند ژست و هویت را کنترل کنند، در حالی که برخی دیگر جزئیات دقیق‌تری مانند بافت مو، رنگ چشم و حالت چهره را کنترل می‌کنند. این نمایش گسسته سبک امکان کنترل دقیق بر خروجی تولید شده را فراهم می‌کند.

StyleGANها همچنین تکنیکی به نام تزریق نویز^۲ را معرفی می‌کنند. نویز تصادفی در نقاط مختلف به مولد اضافه می‌شود که به ایجاد تغییرات تصادفی در تصاویر تولید شده کمک می‌کند. این به جلوگیری از به خاطر سپردن داده‌های آموزشی توسط مولد کمک می‌کند و تنوع خروجی تولید شده را بهبود می‌بخشد.

فرآیند آموزش برای StyleGANها مشابه سایر GANها است که شامل آموزش خصمانه بین مولد و یک متمایزکننده است. با این حال، به دلیل پیچیدگی معماری StyleGAN، آموزش می‌تواند محاسباتی فشرده باشد.

StyleGAN به نتایج قابل توجهی در تولید تصاویر فوتورئالیستی از چهره‌ها دست یافته است. آن‌ها همچنین برای سایر وظایف تولید تصویر مانند تولید تصاویر اتومبیل‌ها، مناظر و سایر اشیاء به کار رفته‌اند.

یکی از مهم‌ترین کمک‌های StyleGANها، توانایی دستکاری تصاویر تولید شده با تغییر بردار استایل است. با درون یابی بین بردارهای سبک مختلف، می‌توان به آرامی بین ویژگی‌های چهره، حالت‌ها یا سبک‌های مختلف انتقال داد. این امر فرصت‌های هیجان‌انگیزی را برای ویرایش و دستکاری تصویر باز می‌کند.

StyleGANها همچنین الهام‌بخش تحقیقات بیشتر در مورد مدل‌های تولیدی بوده‌اند و منجر به توسعه معماری‌های پیشرفته‌تر مانند StyleGAN2 و StyleGAN3 شده‌اند که برخی از محدودیت‌های StyleGAN اصلی را برطرف می‌کنند و کیفیت و کنترل‌پذیری تصاویر تولید شده را بهبود می‌بخشند. StyleGANها واقعاً مرزهای سنتز تصویر را جابجا کرده‌اند و قدرت مدولاسیون سبک را برای تولید تصاویر بسیار واقعی و قابل کنترل نشان داده‌اند.

3-6 شبکه Progressive GAN[7]

ProgressiveGanهای (هم‌چنین به عنوان ProGAN شناخته می‌شوند) یک روش آموزشی جدید را معرفی کردند که به طور قابل توجهی پایداری و کیفیت GANها را بهبود بخشید، به ویژه برای تولید تصاویر با وضوح بالا. ایده اصلی پشت ProGANها رشد تدریجی مولد و متمایزکننده در طول آموزش است.

¹ Style vector
² Noise injection

آموزش سنتی GAN اغلب شامل آموزش کل شبکه به طور همزمان است که می‌تواند چالش برانگیز باشد، به خصوص برای تولید تصاویر با وضوح بالا. ProGAN ها رویکرد متفاوتی دارند. آنها با آموزش GAN ها بر روی تصاویر با وضوح بسیار پایین (به عنوان مثال، پیکسل‌های 4×4) شروع می‌کنند. در این مرحله اولیه، شبکه‌ها یاد می‌گیرند که ساختار درخت و ویژگی‌های اساسی داده‌ها را ضبط کنند.

با پیشرفت آموزش، ProGAN ها به تدریج لایه‌های جدیدی را هم به مولد و هم به متمایزکننده اضافه می‌کنند. این لایه‌های جدید وظیفه مدل‌سازی جزئیات دقیق‌تر و افزایش وضوح تصاویر تولید شده را بر عهده دارند. مولد با نمونه‌برداری از نقشه‌های ویژگی با وضوح پایین و متمایزکننده با پایین‌نمونه‌سازی تصاویر ورودی با وضوح بالا شروع می‌کند.

کلید این رویکرد رو به رشد پیش‌رونده این است که شبکه‌ها مجبور نیستند همه مقیاس‌های جزئیات را به طور همزمان یاد بگیرند. در عوض، آنها یاد می‌گیرند که با پیشرفت آموزش، روی جزئیات بیشتر و دقیق‌تر تمرکز کنند. این مسئله یادگیری را آسان‌تر و پایدارتر می‌کند و به GAN ها اجازه می‌دهد تا تصاویر با کیفیت بالاتر و جزئیات بیشتری تولید کنند.

وقتی لایه‌های جدید اضافه می‌شوند، به آرامی "محو" می‌شوند. این بدان معناست که خروجی‌های لایه‌های قدیمی و جدید با استفاده از میانگین وزنی ترکیب می‌شوند و با ادامه آموزش به تدریج تمرکز به لایه‌های جدید تغییر می‌کند. این انتقال آرام کمک می‌کند تا از ضربه‌های ناگهانی به لایه‌هایی که از قبل به خوبی آموزش دیده‌اند، با وضوح پایین‌تر جلوگیری شود.

ProGANs همچنین چند تکنیک دیگر را برای بهبود بیشتر ثبات آموزش و کیفیت تصویر معرفی کرد:

- **انحراف استاندارد مینی دسته:** این تکنیک اندازه‌ای از تغییرات درون هر مینی دسته را به ورودی متمایزکننده اضافه می‌کند که به بهبود تنوع نمونه‌های تولید شده کمک می‌کند.

- **نرمال‌سازی پیکسل:** این تکنیک بردارهای ویژگی را در مولد عادی می‌کند، که به تثبیت آموزش و جلوگیری از فروپاشی حالت کمک می‌کند.

- **نرخ یادگیری برابر^۱:** این تکنیک وزن‌های شبکه را در زمان اجرا مقیاس می‌کند، که به جلوگیری از تشدید بزرگی سیگنال و بهبود پایداری آموزش کمک می‌کند.

ProGAN ها در تولید تصاویر با وضوح بالا از چهره‌ها، اتاق خواب‌ها و سایر اشیاء بسیار موفق بوده‌اند. آنها اثربخشی رشد تدریجی را برای تثبیت آموزش GAN و بهبود کیفیت تصویر نشان داده‌اند.

¹Faded in
²Equalized Learning Rate

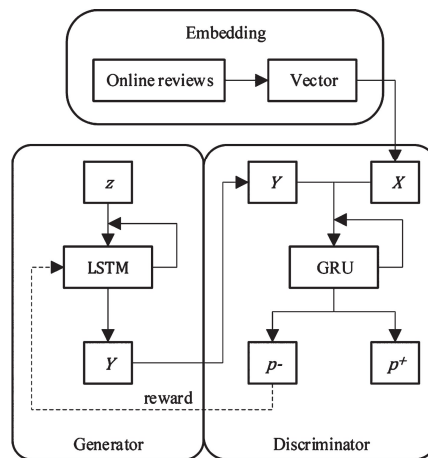
ProGAN ها همچنین الهام بخش تحقیقات بیشتر در مورد روش های آموزشی چند مقیاسی برای GAN ها بوده اند و راه را برای معماری های پیشرفته تر مانند StyleGAN ها هموار کرده اند که بر اساس ایده های معرفی شده در ProGAN ها ساخته شده اند.

3-7 شبکه متخاصم سازنده متوالی

SeqGAN (شبکه متخاصم سازنده متوالی¹) یک نوع تخصصی از GAN است که برای تولید دنباله هایی از توکن های مجزا مانند متن، موسیقی یا نمایش نمادین داده های دیگر طراحی شده است. GAN های سنتی، که در فضاهای داده پیوسته کار می کنند، به طور مستقیم برای تولید دنباله قابل استفاده نیستند زیرا فضای خروجی گسسته و غیر قابل تمایز است. SeqGAN ها این چالش را با ترکیب آموزش خصمانه با یادگیری تقویتی برطرف می کنند.

ایده کلیدی پشت SeqGAN ها استفاده از متمایز کننده به عنوان یک تابع پاداش برای عامل یادگیری تقویتی است. مولد به عنوان عامل عمل می کند و هدف آن تولید دنباله هایی است که پاداش داده شده توسط متمایز کننده را به حداکثر می رساند. متمایز کننده، طبق معمول، سعی می کند بین دنباله های واقعی و تولید شده تمایز قائل شود.

در این شبکه ها مولد یک شبکه عصبی بازگشتی (RNN) است که معمولاً یک LSTM یا GRU است که به عنوان یک شبکه خط مشی در یادگیری تقویتی عمل می کند. یک بردار پنهان (نویز) را به عنوان ورودی می گیرد و دنباله ای از توکن های گسسته، یک توکن را در یک زمان تولید می کند. پارامترهای مولد برای به حداکثر رساندن پاداش مورد انتظار از متمایز کننده آموخته می شوند.



شکل 3-7 ساختار شبکه seq gan

¹ Sequence Generative Adversarial Network

متمایز کننده نیز یک شبکه عصبی بازگشتی است که دنباله‌ای از توکن‌ها را به عنوان ورودی می‌گیرد و احتمال واقعی بودن دنباله را تعیین می‌کند.

فرآیند آموزش شامل یادگیری تقویتی است. مولد با استفاده از گرادینت‌ها آموزش داده می‌شود تا پاداش مورد انتظار از متمایز کننده را به حداکثر برساند. پاداش توسط خروجی متمایز کننده داده می‌شود، که نشان می‌دهد دنباله تولید شده چقدر واقعی است. مولد برای تولید دنباله‌هایی که متمایز کننده به عنوان واقعی طبقه‌بندی می‌کند، پاداش می‌گیرد.

متمایز کننده با استفاده از تکنیک‌های آموزش متخاصم استاندارد آموزش می‌بیند. هم دنباله‌های واقعی را از داده‌های آموزشی دریافت می‌کند و هم دنباله‌های جعلی تولید شده توسط مولد را دریافت می‌کند. هدف آن تشخیص دنباله‌های واقعی و جعلی است.

از آنجایی که سیگنال پاداش از متمایز کننده فقط برای دنباله‌های کامل در دسترس است، عرضه مونت کارلو برای تخمین پاداش برای دنباله‌های جزئی تولید شده توسط مولد استفاده می‌شود. این شامل نمونه‌برداری از دنباله‌های کامل چندگانه از مولد، شروع از دنباله جزئی، و سپس میانگین گرفتن پاداش از متمایز کننده برای این دنباله کامل است.

فرآیند آموزش به طور متناوب بین به روز رسانی مولد با استفاده از گرادینت‌ها و به روز رسانی متمایز کننده با استفاده از آموزش خصمانه است. این روند تا زمانی ادامه می‌یابد که مولد یاد بگیرد دنباله‌هایی تولید کند که بتواند متمایز کننده را فریب دهد و متمایز کننده در شناسایی دنباله‌های جعلی بهتر می‌شود.

کاربردهای معمول	محدودیت	نقاط قوت	مشکلات	معماری کلیدی	تابع ضرر	ورودی	ورودی مولد	ایده اصلی	Feature
تولید تصویر ساده	بی ثباتی آموزش، فروپاشی حالت	مدل پایه GAN	ناپایداری، فروپاشی حالت، کاهش شیب	شبکه‌های عصبی ساده	مینی ماکس یا غیر اشباع	داده‌های واقعی x / داده‌های جعلی $G(z)$	نویز تصادفی (z)	آموزش خصمانه (مولد در مقابل متمایز کننده)	وانیل
تولید تصویر با برچسب، ویرایش تصویر	به داده‌های دارای برچسب اشراط نیاز دارد	نسل کنترل شده	مانند وانیل GAN، به علاوه مشکلات	ورودی مشروط به G و D	Minimax شرطی یا غیر اشباع	داده‌های واقعی x / (داده‌های جعلی $G(z)$ ، y)	$(z) +$ اطلاعات شرطی (y)	تولید مشروط بر برچسب‌ها/اطل	CGAN
تولید تصویر با وضوح بالا	هزوز هم می تواند از بی ثباتی آموزش رنج ببرد	کیفیت تصویر به طور قابل توجهی بهبود یافته است	CNNهای آموزشی، فروپاشی حالت، ناپدید	لایه‌های کانولوشن در D و G هر دو	Minimax اصلاح شده یا غیر اشباع با معماری CNN	داده‌های واقعی x / داده‌های جعلی $G(z)$	نویز تصادفی (z)	استفاده از CNN برای تولید تصویر	DCGAN
تولید تصاویر متنوع و با کیفیت	هزینه محاسباتی، تنظیم فرایا کمتر	آموزش پایدارتر، متریک	محدودیت Lipschitz، برش وزن /	شبکه منتقد به جای متمایز کننده	تقریب فاصله Wasserstein	داده‌های واقعی x / داده‌های جعلی $G(z)$	نویز تصادفی (z)	فاصله Wasserstein برای عملکرد ضرر	WGAN
ترجمه تصویر به تصویر بدون جفت	آموزش موثر می تواند چالش برانگیز باشد	ترجمه تصویر به تصویر جفت نشده	ثبات آموزش، تنظیم ضرر	دو مولد G و F دو متمایز کننده	ترکیبی از خصوصیت و ثبات چرخه	داده‌های واقعی $x(y)$ / داده‌های جعلی $G(z/F(y))$	نویز تصادفی (z)	ثبات چرخه برای ترجمه تصویر جفت	CycleGAN
تولید چهره با وضوح بالا، دستکاری	هزینه محاسباتی، بالقوه برای مصنوعات	تصاویر با وضوح بالا، واقعی، کنترل	هزینه محاسباتی، فروپاشی حالت	شبکه نقشه برداری، رشد تدریجی، مدولاسیون سبک	زبان خصمانه اصلاح شده با مدولاسیون	داده‌های واقعی x / داده‌های جعلی $G(z)$	نویز $(z) +$ نگاشت شبکه به بردار سبک (s)	مدولاسیون سبک برای کنترل ریزدانه	StyleGAN
تولید تصویر با وضوح بالا	معماری پیچیده، می تواند محاسباتی فشرده باشد	آموزش پایدار، تصاویر با وضوح بالا	ثبات آموزش در وضوح بالا	رشد پیشرونده لایه‌های G و D	رشد خصمانه اصلاح شده با رشد تدریجی	داده‌های واقعی x / داده‌های جعلی $G(z)$	نویز تصادفی (z)	رشد پیشرونده برای تصاویر با وضوح بالا	ProGAN

فصل 4: چالش‌ها و محدودیت‌ها

GANها، با وجود قابلیت‌های چشمگیرشان، آموزش بسیار چالش برانگیزی هستند. آن‌ها مستعد بی‌ثباتی‌های مختلف هستند که می‌تواند مانع از فرآیند یادگیری شود و منجر به نتایج غیربهبوده شود. دو تا از برجسته ترین چالش‌ها فروپاشی حالت و ناپدید شدن گرادیان‌ها هستند.

4-1 بی‌ثباتی آموزش

4-1-1 فروپاشی حالت

فروپاشی حالت پدیده‌ای است که در آن مولد گیر می‌کند و تنها تعداد محدودی از خروجی‌ها را تولید می‌کند، حتی اگر توزیع واقعی داده‌ها بسیار متنوع‌تر باشد. یک GAN را تصور کنید که بر روی تصاویر ارقام دست نویس آموزش دیده است. به جای تولید هر ده رقم (0-9)، ممکن است به طور مداوم فقط رقم "7" را بدون توجه به نویز ورودی ایجاد کند. مولد اساساً بر روی یک "حالت" از توزیع داده سقوط کرده است و حالت‌های دیگر را نادیده می‌گیرد.



شکل 4-1 نمونه تصویر ساخته‌شده توسط شبکه در فروپاشی حالت

فروپاشی حالت مشکل ساز است زیرا تنوع و سودمندی داده‌های تولید شده را محدود می‌کند. این نشان می‌دهد که مولد توزیع داده‌های اساسی را به طور کامل یاد نگرفته است و قادر به گرفتن پیچیدگی کامل آن نیست. چندین دلیل بالقوه برای فروپاشی حالت وجود دارد، از جمله:

- **مولد و ممیز متعادل:** اگر مولد در مقایسه با متمایزکننده خیلی ضعیف باشد، ممکن است به جای تلاش برای گرفتن تنوع کامل داده‌ها، تخصص روی یک حالت واحد که بتواند متمایزکننده را فریب دهد، آسان تر است.
- **مشکلات عملکرد ضرر:** برخی از توابع ضرر یا استراتژی‌های آموزشی می‌توانند بیشتر از سایرین مستعد سقوط حالت باشند.
- **فقدان داده کافی:** اگر داده‌های آموزشی محدود باشد، مولد ممکن است بیش از حد مناسب باشد و زیرمجموعه کوچکی از داده‌ها را به خاطر بسپارد و منجر به فروپاشی حالت شود.

4-1-2 ناپدیدشدن گرادیان

ناپدیدشدن گرادیان زمانی اتفاق می‌افتد که شیب تابع ضرر با توجه به پارامترهای مولد در طول آموزش بسیار کوچک شود. این امر یادگیری و بهبود را برای مولد دشوار می‌کند زیرا به‌روزرسانی‌های پارامترهای آن ناچیز و بی‌اثر هستند. تصور کنید سعی می‌کنید یک جعبه سنگین را فشار دهید، اما دستان شما مدام می‌لغزند – نمی‌توانید آن را زیاد حرکت دهید.

ناپدیدشدن گرادیان به ویژه در مراحل اولیه آموزش GAN که مولد ضعیف است مشکل ساز است. اگر متمایزکننده بسیار قوی‌تر از مولد باشد، به راحتی می‌تواند داده‌های واقعی را از جعلی تشخیص دهد. این می‌تواند باعث شود که شیب‌هایی که به سمت مولد برمی‌گردند بسیار کوچک شوند و توانایی آن برای یادگیری را مختل کند. مولد به طور موثر "گیر" می‌کند زیرا بازخورد مفید کافی از متمایزکننده دریافت نمی‌کند.

شیب ناپدید شدن می‌تواند ناشی از موارد زیر باشد:

- **خیلی سریع قوی شدن متمایزکننده:** اگر متمایزکننده خیلی سریع یاد بگیرد، می‌تواند کار را برای مولد دشوار کند.
- **انتخاب تابع ضرر:** برخی از توابع ضرر بیشتر از بقیه مستعد ناپدیدشدن گرادیان هستند.

4-1-3 جمع‌بندی

فروپاشی حالت و ناپدیدشدن گرادیان چالش‌های درهم تنیده هستند. پرداختن به یکی گاهی می‌تواند دیگری را تشدید کند. به عنوان مثال، تلاش برای جلوگیری از فروپاشی حالت با قوی‌تر کردن مولد ممکن است منجر به ناپدیدشدن گرادیان‌ها شود، اگر متمایزکننده نتواند ادامه دهد.

محققان تکنیک‌های مختلفی را برای کاهش این مشکلات توسعه داده اند، از جمله:

- **توابع ضرر بهبودیافته:** برای مثال WGAN ها با استفاده از فاصله Wasserstein به مشکل ناپدید شدن گرادیان رسیدگی می کنند.
- **معماری‌های مختلف شبکه:** به عنوان مثال، DCGAN ها ثبات بهبود یافته‌ای را برای تولید تصویر نشان داده‌اند.
- **تنظیم دقیق‌های پیرامونتر:** نرخ یادگیری، اندازه دسته و سایر پارامترها می توانند به طور قابل توجهی بر ثبات آموزش تأثیر بگذارند.
- **تکنیک‌های منظم سازی:** تکنیک‌هایی مانند ترک تحصیل و کاهش وزن می توانند به جلوگیری از تناسب بیش از حد و بهبود تعمیم کمک کنند.

علی‌رغم این پیشرفت‌ها، GAN های آموزشی یک عمل متعادل کننده ظریف باقی می ماند و این ناپایداری‌های آموزشی هنوز حوزه‌های فعال تحقیقاتی هستند. یافتن استراتژی‌های آموزشی قوی و قابل اعتماد برای GAN ها یک تلاش مداوم در زمینه یادگیری عمیق است.

فصل 5: معیارهای ارزیابی

یکی از چالش‌های اصلی در تحقیقات GAN، ارزیابی عینی کیفیت و تنوع نمونه‌های تولید شده است. بر خلاف وظایف یادگیری نظارت شده که در آن‌ها برچسب‌های حقیقت پایه^۱ برای مقایسه داریم، در GAN‌ها، ما در تلاش برای مدل سازی یک توزیع داده هستیم که ذاتاً ارزیابی آن دشوارتر است. چگونه بفهمیم که یک GAN تصاویر "خوب" تولید می‌کند؟

این جایی است که معیارهای ارزیابی وارد می‌شوند. آن‌ها معیارهای کمی را برای ارزیابی جنبه‌های مختلف عملکرد GAN ارائه می‌دهند. دو مورد از پرکاربردترین معیارها عبارتند از: امتیاز شروع^۲ (IS) و فاصله شروع فریشت^۳ (FID).

5-1 امتیاز اولیه (IS): اندازه گیری کیفیت و تنوع

هدف Inception Score گرفتن دو ویژگی کلیدی از تصاویر تولید شده خوب است:

- **کیفیت تصویر:** تصاویر تولید شده باید واقعی به نظر برسند و به راحتی قابل طبقه‌بندی باشند.
- **تنوع تصویر:** تصاویر تولید شده باید طیف متنوعی از دسته‌ها یا ویژگی‌ها را در بر گیرند.

برای محاسبه IS، از یک مدل Inception از پیش آموزش دیده استفاده می‌کنیم، که یک شبکه عصبی کانولوشنال عمیق است که بر روی یک مجموعه داده تصویری عظیم (مانند ImageNet) آموزش داده شده است. مدل Inception برای طبقه بندی تصاویر تولید شده استفاده می‌شود.

IS به صورت زیر محاسبه می‌شود:

$$IS = \exp(E_x[kL(p(y|x) || p(y))])$$

که در آن:

- x تصویر تولید شده
- $p(y|x)$ توزیع کلاس شرطی پیش بینی شده توسط مدل Inception برای تصویر x
- $p(y)$ توزیع کلاس حاشیه‌ای تصاویر تولید شده
- KL نشان دهنده واگرایی Kullback-Leibler

¹ Ground truth

² Inception Score

³ Fréchet Inception Distance

IS بالاتر نشان‌دهنده عملکرد بهتر است. یک GAN خوب باید تصاویری تولید کند که هم به راحتی قابل طبقه‌بندی ($p(y|x)$ بالا) و هم متنوع (آنتروپی بالا در $p(y)$) باشند.

با این حال، IS محدودیت‌هایی دارد. این مدل به مدل Inception متکی است که بر روی یک مجموعه داده خاص (ImageNet) آموزش داده شده است. این می‌تواند سوگیری ایجاد کند و ممکن است برای ارزیابی GAN‌های آموزش دیده بر روی انواع بسیار متفاوت داده مناسب نباشد. همچنین، IS در درجه اول به کیفیت تصاویر فردی حساس است و ممکن است تنوع کلی داده‌های تولید شده را به طور کامل ثبت نکند.

5-2 فاصله شروع فریشت (FID): اندازه‌گیری قوی‌تر

فاصله اولیه Fréchet به برخی از محدودیت‌های IS می‌پردازد. توزیع تصاویر واقعی و تولید شده را در فضای ویژگی مدل Inception مقایسه می‌کند. FID به جای اینکه فقط به تصاویر منفرد نگاه کند، ویژگی‌های آماری کل مجموعه تصاویر تولید شده را در نظر می‌گیرد.

FID به صورت زیر محاسبه می‌شود:

$$F = \|\mu_r - \mu_g\|^2 + \text{Tr}(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g))^{1/2}$$

که در آن:

- μ_r و μ_g به ترتیب بردارهای مشخصه میانگین تصاویر واقعی و تولید شده
- Σ_r و Σ_g به ترتیب ماتریس‌های کوواریانس تصاویر واقعی و تولید شده
- Tr نشان دهنده trace ماتریس است.¹

FID پایین‌تر نشان‌دهنده عملکرد بهتر است. یک GAN خوب باید تصاویری تولید کند که توزیع ویژگی آن‌ها با داده‌های واقعی مطابقت داشته باشد.

FID به طور کلی یک معیار قوی‌تر و قابل اعتمادتر از IS در نظر گرفته می‌شود. حساسیت کمتری نسبت به نویز و نقاط پرت دارد و معیار بهتری از کیفیت کلی و تنوع داده‌های تولید شده ارائه می‌دهد.

IS و FID هر دو ابزار ارزشمندی برای ارزیابی GAN‌ها هستند، اما کامل نیستند. آن‌ها به مدل Inception متکی هستند و ممکن است برای همه انواع داده‌ها مناسب نباشند. محققان به طور مستمر در حال بررسی معیارهای ارزیابی جدید و بهبود یافته برای GAN‌ها هستند تا تفاوت‌های ظریف مدل سازی مولد را بهتر درک کنند. در نهایت، ترکیبی از معیارهای کمی و ارزیابی کیفی (بازرسی بصری) اغلب برای ارزیابی عملکرد GAN استفاده می‌شود.

¹ جمع درایه‌های قطر اصلی

3-5 الزامات داده و هزینه محاسباتی: قیمت واقع گرایی

GANها، در حالی که قادر به تولید داده‌های واقعی خیره‌کننده هستند، داده‌ها و نیازهای محاسباتی قابل توجهی دارند. این الزامات می‌تواند یک مانع بزرگ برای محققان یا متخصصان با منابع محدود باشد.

1-3-5 نیازهای داده

GANها معمولاً به مقادیر زیادی از داده‌های آموزشی با کیفیت بالا برای یادگیری مؤثر توزیع داده‌های پیچیده نیاز دارند. هرچه داده‌ها پیچیده‌تر باشند (مانند تصاویر با وضوح بالا، فیلم‌ها)، GAN به طور کلی به داده‌های بیشتری نیاز دارد. داده‌های ناکافی می‌تواند منجر به بیش‌برازش شود، جایی که GAN داده‌های آموزشی را به خاطر می‌سپارد و نمی‌تواند به داده‌های جدید و نادیده گرفته و تعمیم یابد. این می‌تواند به صورت کیفیت پایین تصویر، فروپاشی حالت، یا سایر مصنوعات نامطلوب ظاهر شود.

به دست آوردن و مدیریت مجموعه داده‌های بزرگ و تمیز می‌تواند به خودی خود چالش مهمی باشد. جمع‌آوری داده‌ها می‌تواند پرهزینه و وقت‌گیر باشد و تمیز کردن داده‌ها (حذف نویز، تصحیح خطاها) می‌تواند یک فرآیند خسته‌کننده باشد. در برخی از حوزه‌ها، مانند تصویربرداری پزشکی، داده‌های برچسب‌گذاری شده ممکن است کمیاب باشند، که آموزش مؤثر GANها را دشوار می‌کند.

2-3-5 هزینه محاسباتی: تلاشی با منابع فشرده

آموزش GANها از نظر محاسباتی گران است. GANها شبکه‌های عصبی عمیق با میلیون‌ها پارامتر هستند و آموزش آن‌ها به قدرت محاسباتی قابل توجهی نیاز دارد. آموزش GAN از ابتدا می‌تواند روزها، هفته‌ها یا حتی ماه‌ها در GPUهای قدرتمند طول بکشد. این باعث می‌شود که آموزش GAN به یک تلاش منابع فشرده تبدیل شود، که اغلب نیاز به دسترسی به خوشه‌های GPU یا خدمات رایانش ابری دارد.

هزینه محاسباتی آموزش GAN به عوامل مختلفی بستگی دارد، از جمله:

- **معماری شبکه:** شبکه‌های عمیق‌تر و گسترده‌تر معمولاً به محاسبات بیشتری نیاز دارند.
- **اندازه و پیچیدگی داده‌ها:** مجموعه داده‌های بزرگ‌تر و پیچیده‌تر به محاسبات بیشتری نیاز دارند.
- **مدت زمان آموزش:** آموزش برای دوره‌های بیشتر به طور کلی نیاز به محاسبات بیشتری دارد.
- **سخت‌افزار:** پردازنده‌های گرافیکی قوی‌تر می‌توانند سرعت آموزش را افزایش دهند، اما هزینه بیشتری نیز دارند.

5-3-3 کاهش چالش‌ها

با وجود این چالش‌ها، محققان به طور فعال در حال توسعه روش‌هایی برای کاهش داده‌ها و نیازهای محاسباتی GAN‌ها هستند. برخی از رویکردها عبارتند از:

- **افزایش داده‌ها:** تکنیک‌هایی برای افزایش مصنوعی اندازه داده‌های آموزشی با اعمال تبدیل به داده‌های موجود.
- **یادگیری انتقالی:** استفاده از GAN‌های از پیش آموزش دیده در وظایف مرتبط برای کاهش میزان داده‌های آموزشی مورد نیاز.
- **فشرده سازی مدل:** روش‌هایی برای کاهش اندازه و پیچیدگی مدل‌های GAN بدون به خطر انداختن عملکرد.
- **آموزش توزیع شده:** استفاده از چندین GPU یا ماشین برای موازی کردن فرآیند آموزش و کاهش زمان آموزش.
- **معماری‌های کارآمد:** طراحی معماری‌های GAN کارآمدتر که نیاز به محاسبات کمتری دارند.

در حالی که این تکنیک‌ها می‌توانند به کاهش داده‌ها و بار محاسباتی GAN کمک کنند، این چالش‌ها همچنان قابل توجه هستند و تحقیقات در حال انجام برای دسترسی بیشتر و کاربردی‌تر کردن GAN‌ها برای طیف وسیع‌تری از کاربردها بسیار مهم است.

5-4 نگرانی‌های اخلاقی و سوء استفاده بالقوه: سمت تاریک خلقت

GAN‌ها با توانایی خود در تولید داده‌های مصنوعی فوق‌العاده واقعی، نگرانی‌های اخلاقی زیادی را ایجاد می‌کنند و فرصت‌هایی را برای سوءاستفاده بالقوه ارائه می‌دهند. این نگرانی‌ها دامنه‌های مختلفی را شامل می‌شود، از اطلاعات نادرست و حریم خصوصی گرفته تا حق چاپ و سوگیری.

5-4-1 دیپ فیک و اطلاعات غلط: فرسایش اعتماد

یکی از نگران‌کننده‌ترین نگرانی‌های اخلاقی پیرامون GAN‌ها، ایجاد دیپ‌فیک است. دیپ‌فیک‌ها رسانه‌های مصنوعی (تصاویر، ویدیوها، صداها) هستند که به گونه‌ای دستکاری می‌شوند که به طور قانع‌کننده‌ای، شخصی را در حال گفتن یا انجام کاری که واقعاً انجام نداده است، به تصویر بکشند. واقع‌گرایی دیپ‌فیک‌ها به جایی رسیده است که تشخیص آن‌ها از رسانه‌های واقعی بسیار دشوار است.

دیپفیک‌ها تهدید قابل توجهی برای انتشار اطلاعات نادرست و از بین رفتن اعتماد به رسانه‌ها هستند. از آن‌ها می‌توان برای ایجاد اخبار جعلی، گسترش تبلیغات، دستکاری افکار عمومی یا حتی آسیب رساندن به شهرت استفاده کرد. پتانسیل استفاده مخرب از دیپفیک بسیار زیاد است و پیامدهای جدی برای جامعه دارد.

5-4-2 نگرانی‌های حفظ حریم خصوصی: محو شدن مرزها

از GAN‌ها می‌توان برای تولید داده‌های مصنوعی که بسیار شبیه به داده‌های واقعی، از جمله اطلاعات شخصی است، استفاده کرد. این نگرانی‌هایی را در مورد حریم خصوصی ایجاد می‌کند، زیرا ممکن است بتوان اطلاعات حساس را از داده‌های مصنوعی بازسازی کرد یا نام افراد را بی‌نام کرد. استفاده از GAN برای تولید داده‌های بیومتریک مصنوعی، مانند چهره یا اثر انگشت، نگرانی‌هایی را در مورد سرقت هویت و تقلب ایجاد می‌کند.

5-4-3 نقض حق چاپ: مسئله مالکیت

GAN‌ها را می‌توان بر روی مطالب دارای حق چاپ، مانند تصاویر، موسیقی یا متن آموزش داد و سپس می‌تواند محتوای جدیدی تولید کند که مشابه مطالب اصلی است. این سؤالاتی را در مورد نقض حق چاپ و حقوق مالکیت معنوی ایجاد می‌کند. چه کسی حق چاپ یک قطعه موسیقی تولید شده توسط GAN آموزش دیده بر روی آهنگ‌های موجود را دارد؟ اینها سؤالات حقوقی و اخلاقی پیچیده‌ای هستند که باید مورد توجه قرار گیرند.

5-4-4 تقویت بایاس: اتاق پژواک هوش مصنوعی

GAN‌ها، مانند سایر مدل‌های یادگیری ماشینی، می‌توانند سوگیری‌های موجود در داده‌های آموزشی خود را به ارث برده و حتی تقویت کنند. اگر داده‌های آموزشی حاوی سوگیری‌های مربوط به جنسیت، نژاد یا سایر ویژگی‌های حساس باشد، GAN ممکن است داده‌های مصنوعی تولید کند که این سوگیری‌ها را تداوم یا اغراق می‌کند. این می‌تواند عواقب جدی داشته باشد، به ویژه در برنامه‌هایی مانند تشخیص چهره، که در آن GAN‌های مغرضانه می‌تواند منجر به نتایج تبعیض آمیز شود.

5-4-5 نیاز به توسعه و استقرار مسئولانه

پرداختن به این نگرانی‌های اخلاقی نیازمند رویکردی چند وجهی است. توسعه و استقرار GAN‌ها به طور مسئولانه و با در نظر گرفتن دقیق تأثیر بالقوه آنها بر جامعه بسیار مهم است. برخی از مراحل کلیدی عبارتند از:

- **توسعه تکنیک‌های تشخیص:** محققان در حال کار بر روی روش‌هایی برای شناسایی دیپفیک و سایر اشکال رسانه‌های مصنوعی هستند. این یک مسابقه تسلیحاتی در حال انجام است، زیرا بازیگران بدخواه دائماً سعی می‌کنند جعلی‌های متقاعدکننده‌تری ایجاد کنند.

- **ارتقاء سواد رسانه ای:** آموزش عمومی در مورد وجود و خطرات احتمالی دیپ فیک بسیار مهم است. مردم باید بدانند که رسانه‌های مصنوعی قابل دستکاری هستند و باید نسبت به اطلاعاتی که مصرف می‌کنند انتقاد کنند.
 - **ایجاد دستورالعمل‌های اخلاقی:** تدوین دستورالعمل‌های اخلاقی برای توسعه و استفاده از GANها ضروری است. این شامل در نظر گرفتن مسائلی مانند حریم خصوصی، حق چاپ و سوگیری است.
 - **مقررات و قوانین:** در برخی موارد، مقررات یا قوانین ممکن است برای رسیدگی به سوء استفاده احتمالی از GAN ضروری باشد.
- GANها یک فناوری قدرتمند با پتانسیل ایجاد مزایای بسیاری هستند. با این حال، آگاهی از پیامدهای اخلاقی آنها و اتخاذ اقداماتی برای کاهش خطرات ناشی از سوءاستفاده از آنها بسیار مهم است. فقط از طریق توسعه و استقرار مسئولانه می‌توانیم اطمینان حاصل کنیم که GANها به نفع استفاده می‌شوند و نه برای آسیب.

فصل 6: کاربردهای GAN در جزئیات

GANها طیف گسترده‌ای از کاربردها را در حوزه‌های مختلف پیدا کرده‌اند، اما تأثیر آن‌ها به ویژه در زمینه تولید و دستکاری تصویر بسیار مهم بوده است. توانایی آن‌ها در تولید تصاویر واقعی و متنوع، صنایع مختلف را متحول کرده و فرصت‌های خلاقانه جدیدی را به وجود آورده است.

6-1 تولید و دستکاری تصویر: یک انفجار خلاقانه

GANها به ابزاری قدرتمند برای تولید و دستکاری تصویر تبدیل شده‌اند که طیف وسیعی از کاربردهای خلاقانه را ممکن می‌سازد:

- **ایجاد تصاویر واقعی:** GANها می‌توانند تصاویری از اشیاء، افراد، صحنه‌ها و حتی مفاهیم انتزاعی تولید کنند که به طور باورنکردنی واقعی به نظر می‌رسند. این در زمینه‌های مختلفی کاربرد دارد، از جمله:
 - **هنر و طراحی:** از GANها می‌توان برای ایجاد سبک‌های هنری جدید، ایجاد بافت‌ها و الگوهای منحصر به فرد یا حتی طراحی محصولات جدید استفاده کرد.
 - **سرگرمی:** از GANها می‌توان برای ایجاد جلوه‌های ویژه در فیلم‌ها، تولید شخصیت‌های واقعی برای بازی‌های ویدیویی و یا حتی ایجاد دنیای مجازی استفاده کرد.
 - **تجارت الکترونیک:** از GANها می‌توان برای تولید تصاویر محصولات از زوایای مختلف یا در تنظیمات مختلف استفاده کرد و تجربه خرید آنلاین را بهبود بخشید.
- **ترجمه تصویر به تصویر:** GANها می‌توانند تصاویر را از یک دامنه به دامنه دیگر ترجمه کنند. به عنوان مثال، آن‌ها می‌توانند طرح‌ها را به عکس تبدیل کنند، تصاویر در مقیاس خاکستری را به تصاویر رنگی تبدیل کنند، یا حتی تصاویر را از یک زبان به زبان دیگر ترجمه کنند (به عنوان مثال، تبدیل یک جمله به زبان انگلیسی به یک تصویر مربوطه). این برنامه کاربردی در:
 - **عکاسی:** از GANها می‌توان برای افزایش کیفیت عکس‌ها، حذف نویز یا حتی افزودن جلوه‌های هنری استفاده کرد.
 - **تصویربرداری پزشکی:** از GANها می‌توان برای بهبود وضوح تصاویر پزشکی استفاده کرد و تشخیص بیماری‌ها را آسان‌تر می‌کند.
 - **تصاویر ماهواره‌ای:** از GANها می‌توان برای افزایش کیفیت تصاویر ماهواره‌ای استفاده کرد و تحلیل کاربری زمین یا نظارت بر تغییرات محیطی را آسان‌تر می‌کند.

- **ویرایش و دستکاری تصویر:** از GANها می‌توان برای ویرایش و دستکاری تصاویر موجود به روش‌های مختلف استفاده کرد. به عنوان مثال، آن‌ها می‌توانند اشیاء را از یک تصویر اضافه یا حذف کنند، پس زمینه را تغییر دهند یا حتی ویژگی‌های چهره را تغییر دهند. این برنامه کاربردی در:
 - **روتوش عکس:** از GANها می‌توان برای روتوش عکس‌ها، از بین بردن لکه‌ها یا چین و چروک‌ها استفاده کرد.
 - **واقعیت مجازی:** از GANها می‌توان برای ایجاد محیط‌های مجازی واقعی برای برنامه‌های VR استفاده کرد.
 - **مد:** از GANها می‌توان برای امتحان کردن لباس‌ها به صورت مجازی یا تولید طرح‌های مد جدید استفاده کرد.
 - **وضوح فوق العاده^۱ و بهبود تصویر:** از GANها می‌توان برای افزایش وضوح تصاویر استفاده کرد و دیدن جزئیات دقیق‌تر را ممکن می‌کند. همچنین می‌توان از آن‌ها برای بهبود کیفیت تصاویر با حذف نویز یا مصنوعات استفاده کرد. این برنامه کاربردی در:
 - **نظارت:** از GANها می‌توان برای افزایش کیفیت فیلم‌های نظارتی استفاده کرد و شناسایی مظنونان را آسان‌تر می‌کند.
 - **تصویربرداری پزشکی:** از GANها می‌توان برای بهبود وضوح تصاویر پزشکی استفاده کرد و تشخیص بیماری‌ها را آسان‌تر می‌کند.
 - **بازسازی عکس‌های تاریخی:** از GANها می‌توان برای بازیابی عکس‌های قدیمی یا آسیب دیده استفاده کرد و آن‌ها را به شکوه اصلی خود بازگرداند.
 - **تولید داده‌های آموزشی:** GANها می‌توانند برای تولید داده‌های آموزشی مصنوعی برای سایر مدل‌های یادگیری ماشین استفاده شوند. این می‌تواند به ویژه زمانی مفید باشد که داده‌های آموزشی برچسب‌گذاری شده کمیاب یا گران باشد. به عنوان مثال، GANها می‌توانند برای تولید تصاویر مصنوعی از چهره‌ها برای آموزش سیستم‌های تشخیص چهره استفاده شوند.
- کاربردهای GANها در تولید و دستکاری تصویر به طور مداوم در حال گسترش است زیرا محققان تکنیک‌های جدید و ابتکاری را توسعه می‌دهند. GANها نه تنها مرزهای آنچه را که در ترکیب تصویر ممکن است جابجا می‌کنند، بلکه تأثیر قابل توجهی بر صنایع مختلف، از هنر و سرگرمی گرفته تا مراقبت‌های بهداشتی و امنیت دارند.

^۱ Super Resolution

فصل 7: جهت گیری های آینده و روندهای تحقیقاتی

حوزه GAN ها به سرعت در حال تکامل است و محققان دائماً ایده های جدید را بررسی می کنند و مرزهای ممکن را در پیش می گیرند. در حالی که GAN ها در کاربردهای مختلف به موفقیت چشمگیری دست یافته اند، هنوز چالش های باز و مسیرهای تحقیقاتی هیجان انگیز زیادی برای پیگیری وجود دارد.

7-1 بهبود ثبات آموزش: تلاش برای استحکام

بی ثباتی آموزش یکی از مهم ترین موانع در تحقیقات GAN است. فروپاشی حالت، ناپدید شدن گرادیان ها و سایر مشکلات آموزشی می تواند آموزش قابل اعتماد و پیوسته GAN ها را چالش برانگیز کند. تحقیقات آینده احتمالاً بر موارد زیر متمرکز خواهد شد:

- **تکنیک های منظم سازی جدید¹:** توسعه روش های منظم سازی جدید برای جلوگیری از بیش برآش و بهبود تعمیم.
- **استراتژی های آموزشی تطبیقی:** طراحی الگوریتم های آموزشی که می توانند به صورت پویا ابرپارامترها یا معماری های شبکه را در طول آموزش تنظیم کنند تا با شرایط متغیر سازگار شوند.
- **درک نظری:** به دست آوردن درک نظری عمیق تر از پویایی آموزش GAN برای شناسایی علل ریشه ای بی ثباتی و ایجاد راه حل های اصولی تر.
- **توابع خطای بهبود یافته:** کاوش توابع خطای جدید که کمتر مستعد ناپایداری های آموزشی هستند و گرادیان های پایدارتری ارائه می دهند.

7-2 افزایش کیفیت و تنوع تصویر: جستجوی کمال

در حالی که GAN ها پیشرفت فوق العاده ای در تولید تصاویر واقعی داشته اند، هنوز هم از نظر کیفیت و تنوع تصویر جا برای بهبود وجود دارد. مسیرهای تحقیقاتی آتی عبارتند از:

- **تولید تصویر با وضوح بالا:** در حال توسعه تکنیک هایی برای تولید تصاویر حتی با وضوح بالاتر با جزئیات ریز.
- **تولید قابل کنترل:** بهبود کنترل بر فرآیند تولید، به کاربران این امکان را می دهد که دقیقاً آنچه را که می خواهند تولید کنند مشخص کنند.
- **تولید نمونه های متنوع:** توسعه روش هایی برای تولید نمونه های متنوع تر که غنای کامل توزیع داده ها را به تصویر می کشد.

¹ Novel regularization technique

- ساخت ویدیوهای سه بعدی: گسترش GAN ها برای تولید مدل ها و ویدئوهای سه بعدی که فرمت های داده پیچیده تری هستند.

7-3 کاوش در معماری های جدید و عملکردهای زیان: موتور نوآوری

جستجو برای معماری های GAN بهتر و توابع ضرر، یک حوزه تحقیقاتی مداوم است. کار آینده احتمالاً شامل موارد زیر خواهد بود:

- مکانیسم های توجه¹: ترکیب مکانیسم های توجه در GAN ها به منظور تمرکز بر بخش های مربوطه از داده های ورودی.
- شبکه های عصبی گراف: بررسی استفاده از شبکه های عصبی گراف برای تولید داده های ساختار یافته نمودار.
- GAN های مبتنی بر ترانسفورمر: ادغام معماری های ترانسفورمر، که موفقیت زیادی در پردازش زبان طبیعی نشان داده اند، در GAN.
- طراحی ضرر خصمانه: توسعه توابع زیان خصمانه جدید که محدودیت های زبان های موجود را برطرف می کند.

7-4 پرداختن به نگرانی های اخلاقی: ضرورت هوش مصنوعی مسئول

همانطور که GAN ها قدرتمندتر می شوند، رسیدگی به نگرانی های اخلاقی مرتبط با استفاده از آن ها بسیار مهم است. تحقیق و توسعه آینده باید بر موارد زیر متمرکز شود:

- تشخیص دیپ فیک: توسعه روش های قوی برای تشخیص دیپ فیک و سایر اشکال رسانه های مصنوعی.
- GAN های حفظ حریم خصوصی: طراحی GAN هایی که می توانند داده های مصنوعی را بدون به خطر انداختن حریم خصوصی افراد تولید کنند.
- کاهش سوگیری: توسعه تکنیک هایی برای کاهش سوگیری ها در داده های تولید شده توسط GAN.
- GAN های قابل توضیح: بهبود تفسیرپذیری GAN ها برای درک نحوه تصمیم گیری آن ها و شناسایی سوگیری های بالقوه.

اینها تنها بخشی از بسیاری از جهت گیری ها و روندهای تحقیقاتی هیجان انگیز آینده در GAN ها هستند. این زمینه پر جنب و جوش و پویا است و می توان انتظار داشت که در سال های آینده شاهد پیشرفت های بیشتری باشیم. همانطور که GAN ها به تکامل خود ادامه می دهند، پتانسیل ایجاد انقلاب در زمینه های مختلف را دارند، از هنر و

¹Attention mechanisms

سرگرمی گرفته تا مراقبت‌های بهداشتی و اکتشافات علمی. با این حال، پرداختن به پیامدهای اخلاقی این فناوری برای اطمینان از استفاده مسئولانه و به نفع جامعه بسیار مهم است.

فصل 8: نتیجه‌گیری

شبکه‌های متخاصم مولد (GANs) به عنوان یک رویکرد قدرتمند و نوآورانه برای مدل‌سازی مولد ظهور کرده‌اند و روش ایجاد و دستکاری داده‌ها را به‌ویژه در حوزه تصاویر متحول کرده‌اند. از تولید تصاویر واقعی از چهره‌ها تا ترجمه بین حوزه‌های تصویر و حتی ایجاد داده‌های آموزشی مصنوعی، GANها قابلیت‌ها و پتانسیل قابل توجه خود را برای تغییر صنایع مختلف نشان داده‌اند.

ایده اصلی آموزش خصمانه، که در آن دو شبکه عصبی (مولد و متمایزکننده) با یکدیگر رقابت می‌کنند، ثابت کرده است که روشی بسیار موثر برای یادگیری توزیع‌های پیچیده داده است. مولد یاد می‌گیرد که نمونه‌های داده‌ای به‌طور فزاینده‌ای واقعی ایجاد کند، در حالی که متمایزکننده یاد می‌گیرد بین داده‌های واقعی و جعلی تمایز قائل شود. این تعامل پویا باعث بهبود هر دو شبکه می‌شود و در نهایت منجر به تولید داده‌های مصنوعی با کیفیت بالا می‌شود.

توسعه معماری‌های مختلف GAN مانند DCGANها، WGANها، CycleGANها، StyleGANها و GANهای پیشرو، بسیاری از چالش‌های مرتبط با GANهای آموزشی را برطرف کرده و قابلیت‌های آن‌ها را به طور قابل توجهی گسترش داده است. این پیشرفت‌ها امکان تولید تصاویر با وضوح بالاتر، کنترل دقیق‌تر فرآیند تولید و حتی یادگیری از داده‌های جفت نشده را فراهم کرده است.

با این حال، GANها بدون محدودیت نیستند. بی‌ثباتی آموزش، از جمله فروپاشی حالت و ناپدید شدن گردیان، یک چالش مهم باقی مانده است. ارزیابی عملکرد GAN می‌تواند دشوار باشد و توسعه معیارهای ارزیابی قوی یک حوزه تحقیقاتی مداوم است. علاوه بر این، GANها می‌توانند از نظر محاسباتی نیاز داشته باشند و به مقادیر زیادی از داده‌های آموزشی نیاز دارند.

شاید مهم‌تر از همه، پیامدهای اخلاقی GANها را نمی‌توان نادیده گرفت. احتمال سوءاستفاده، از جمله ایجاد دیپ‌فیک، نقض حریم خصوصی، نقض حق نسخه‌برداری، و تقویت سوگیری، مستلزم بررسی دقیق و توسعه مسئولانه است. توسعه تکنیک‌هایی برای شناسایی دیپ‌فیک، حفظ حریم خصوصی، کاهش تعصب و ایجاد دستورالعمل‌های اخلاقی برای استفاده از GANها بسیار مهم است.

با وجود این چالش‌ها، آینده GANها روشن است. تحقیقات مداوم در زمینه بهبود پایداری آموزش، افزایش کیفیت و تنوع تصویر، کاوش در معماری‌های جدید و عملکردهای از ضرر، و رسیدگی به نگرانی‌های اخلاقی، پتانسیل بیشتری را باز می‌کند. همانطور که GANها به تکامل خود ادامه می‌دهند، پتانسیل ایجاد انقلاب در زمینه‌های مختلف

را دارند، از هنر و سرگرمی گرفته تا مراقبت‌های بهداشتی، اکتشافات علمی و فراتر از آن. با پرداختن به چالش‌ها و پذیرش توسعه مسئولانه، می‌توانیم از قدرت GAN‌ها برای ایجاد آینده‌ای بهتر استفاده کنیم.

فصل 9: منابع

- [1] I. J. Goodfellow *et al.*, “Generative Adversarial Networks,” 2014, *arXiv*. doi: 10.48550/ARXIV.1406.2661.
- [2] M. Mirza and S. Osindero, “Conditional Generative Adversarial Nets,” 2014, *arXiv*. doi: 10.48550/ARXIV.1411.1784.
- [3] A. Radford, L. Metz, and S. Chintala, “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks,” 2015, *arXiv*. doi: 10.48550/ARXIV.1511.06434.
- [4] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein GAN,” 2017, *arXiv*. doi: 10.48550/ARXIV.1701.07875.
- [5] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks,” in *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice: IEEE, Oct. 2017, pp. 2242–2251. doi: 10.1109/ICCV.2017.244.
- [6] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, “Analyzing and Improving the Image Quality of StyleGAN,” in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA: IEEE, Jun. 2020, pp. 8107–8116. doi: 10.1109/CVPR42600.2020.00813.
- [7] T. Karras, T. Aila, S. Laine, and J. Lehtinen, “Progressive Growing of GANs for Improved Quality, Stability, and Variation,” 2017, *arXiv*. doi: 10.48550/ARXIV.1710.10196.