# شبکه های مولد تخاصمی

# Generative Adversarial Networks (GAN)


ارائه دهندگان :

علیرضا طباطبائیان

مریم برازنده

حمیدرضا صفری

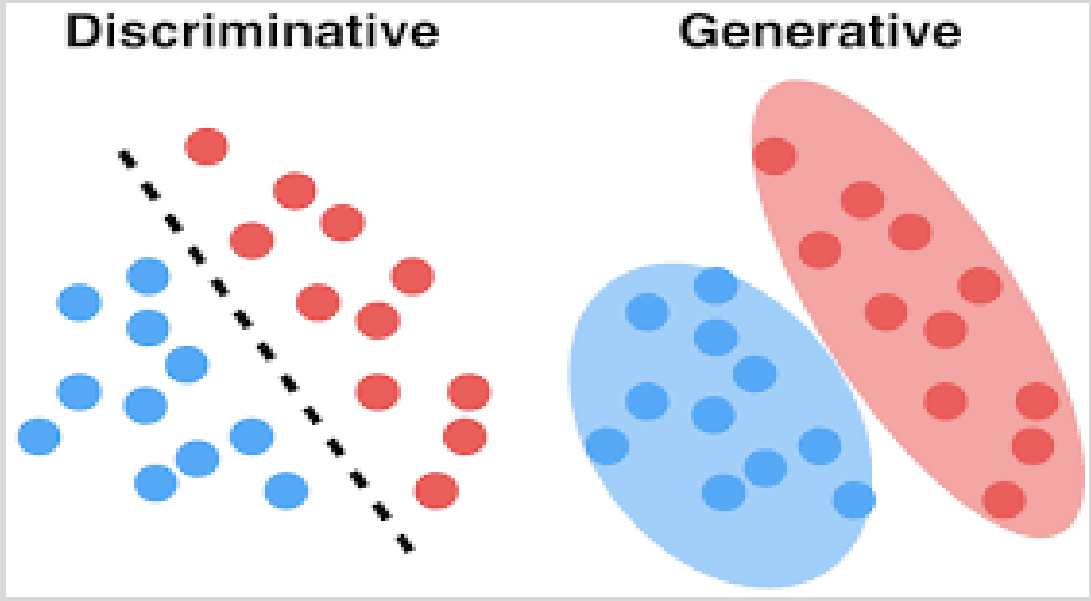
دانش عبداللهی

# موضوعات

- انواع مدل های یادگیری ماشین
- مدل یادگیری تخاصمی چیست ؟
- تابع هزینه
- تولید داده در شبکه های مولد
- کاربرد های GAN
- مشکلات رایج GAN

# انواع مدل های یادگیری ماشین چیست ؟

مدل های یادگیری ماشین به طور کلی به ۲ دسته مدل های مولد یا MODELS GENERATIVE و مدل های متمایزگر یا MODELS DISCRIMINATIVE تقسیم بندی میشوند .



# مدل متمایز گر به چه مدلی میگویند ؟

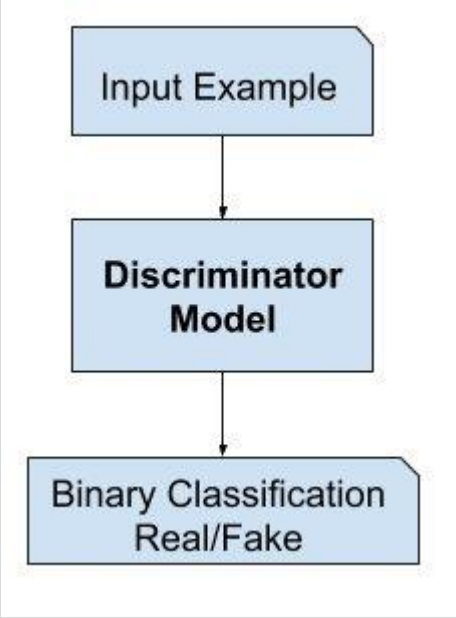
مدل متمایز گر یا Discriminative Modelبه مدلی میگویند که مستقیما تابع هدف که معمولا احتمال شرطی به صورت P(Y|X) است را با کمک دادگان آموزشی پیدا میکند. معمولا از مدل های متمایز گر در طبقه بندی یا Classification استفاده میشود. در این نوع یادگیری ، تمرکز بر تعیین مرز بین داده هاست نه بررسی چگونگی توزیع داده ها. از نمونه مدل های متمایز گر میتوان مدل درخت تصمیم در کاربرد طبقه بندی (Decision Tree) ، مدل جنگل تصادفی (Random Forest) ، مدل شبکه های عصبی (Neural Networks) و مدل نزدیکترین همسایه (Nearest Neighbor) را نام برد.

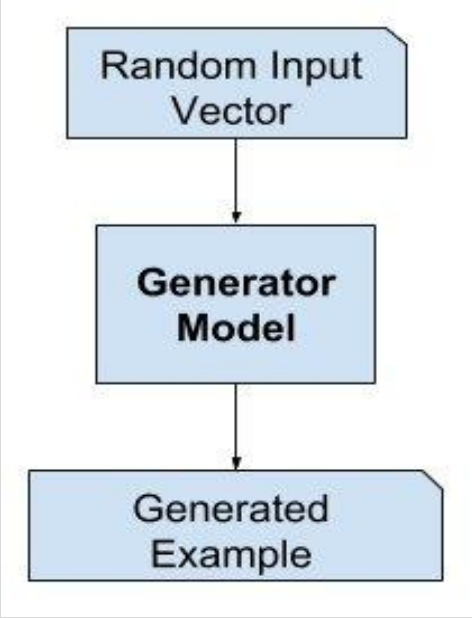
# مدل مولد به چه مدلی میگویند ؟

مدل مولد یا Generative Model به مدلی میگویند که بر روی تابع P(X,Y) کار میکند و معمولا برای یافتن تابع هدف که P(Y|X) باشد ، از رابطه بیز (Bayes) و رابطه زنجیره ای Chain (Rule) و دانش پیشین P(Y) و احتمال شرطی P(X|Y)و سپس ماکسیمم کردن ضرب آنها یعنی argmax(P(X|Y).P(Y))استفاده میکند. از نمونه مدل های مولد میتوان مدل بیز ساده شده (Naive Bayes) ، شبکه های بیزین (Bayesian Networks) ، مدل میدان تصادفی مارکوف (Markov Random Field) و مدل شبکه های مولد تخاصمی (Generative Adversarial Networks(GAN)) که اتفاقا موضوع بحث ما میباشد را نام برد

## مقایسه این دو مدل از نظر کاربرد چگونه است ؟

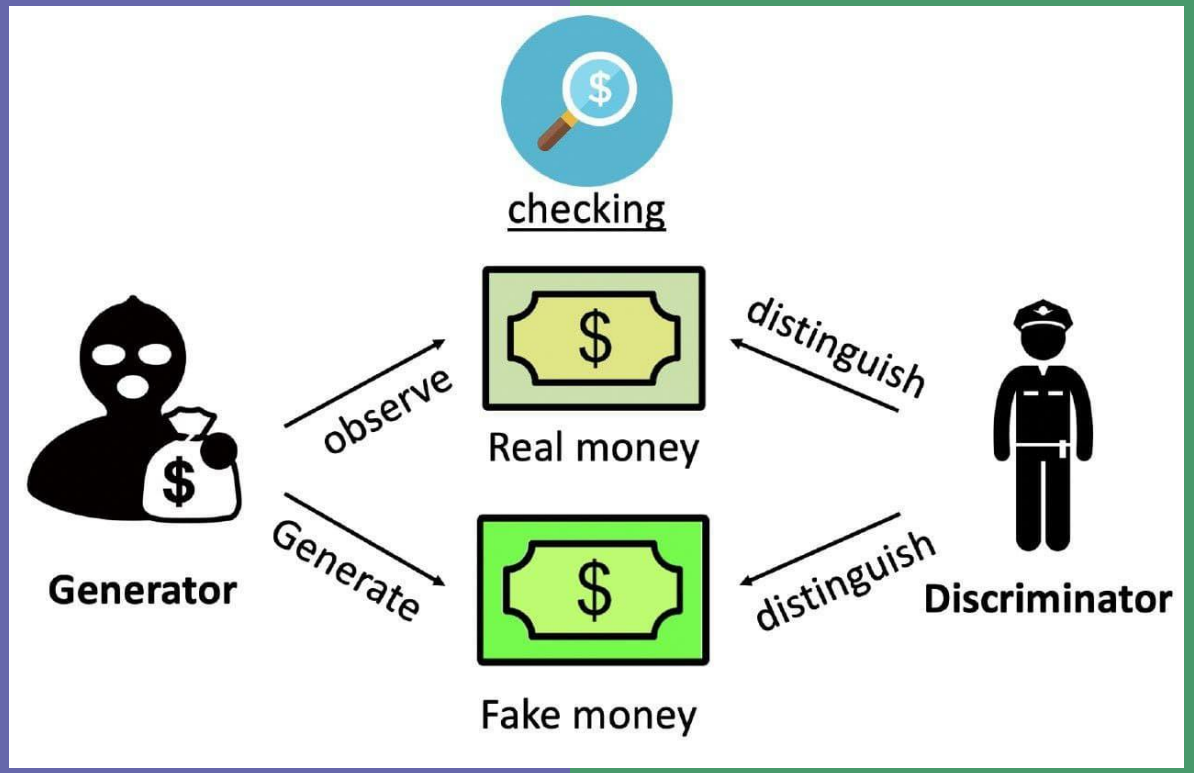
به طور خلاصه میتوان گفت مدل های متمایزگر برای دسته بندی داده ها کاربرد دارند در حالی که مدل های مولد برای تولید داده های جدید کاربرد فراوانی دارند. البته که مدل های مولد ، ریاضیات پیچیده تر و به نسبت، هزینه محاسباتی بیشتری دارند.



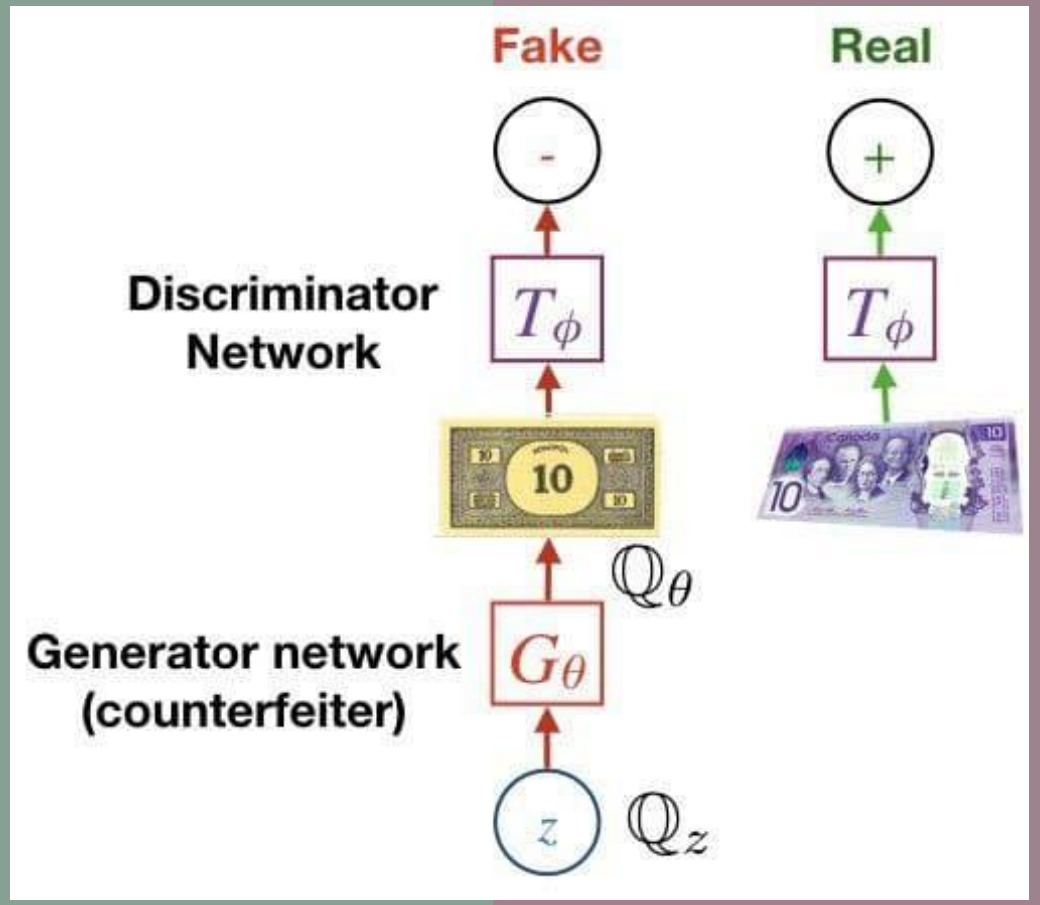
# مدل یادگیری تخاصمی چیست ؟

کلمه تخاصم به معنای دشمنی و ضدیت است. همواره در هر ضدیت ، ۲ شی یا شخص درگیر هستند. طرح اولیه مدل های تخاصمی از همین توصیف الهام گرفته شده است. در این مدل ، ۲ شبکه که اولی یک شبکه Generative و دومی یک شبکه Discriminative میباشد را به صورت همزمان آموزش میدهیم.



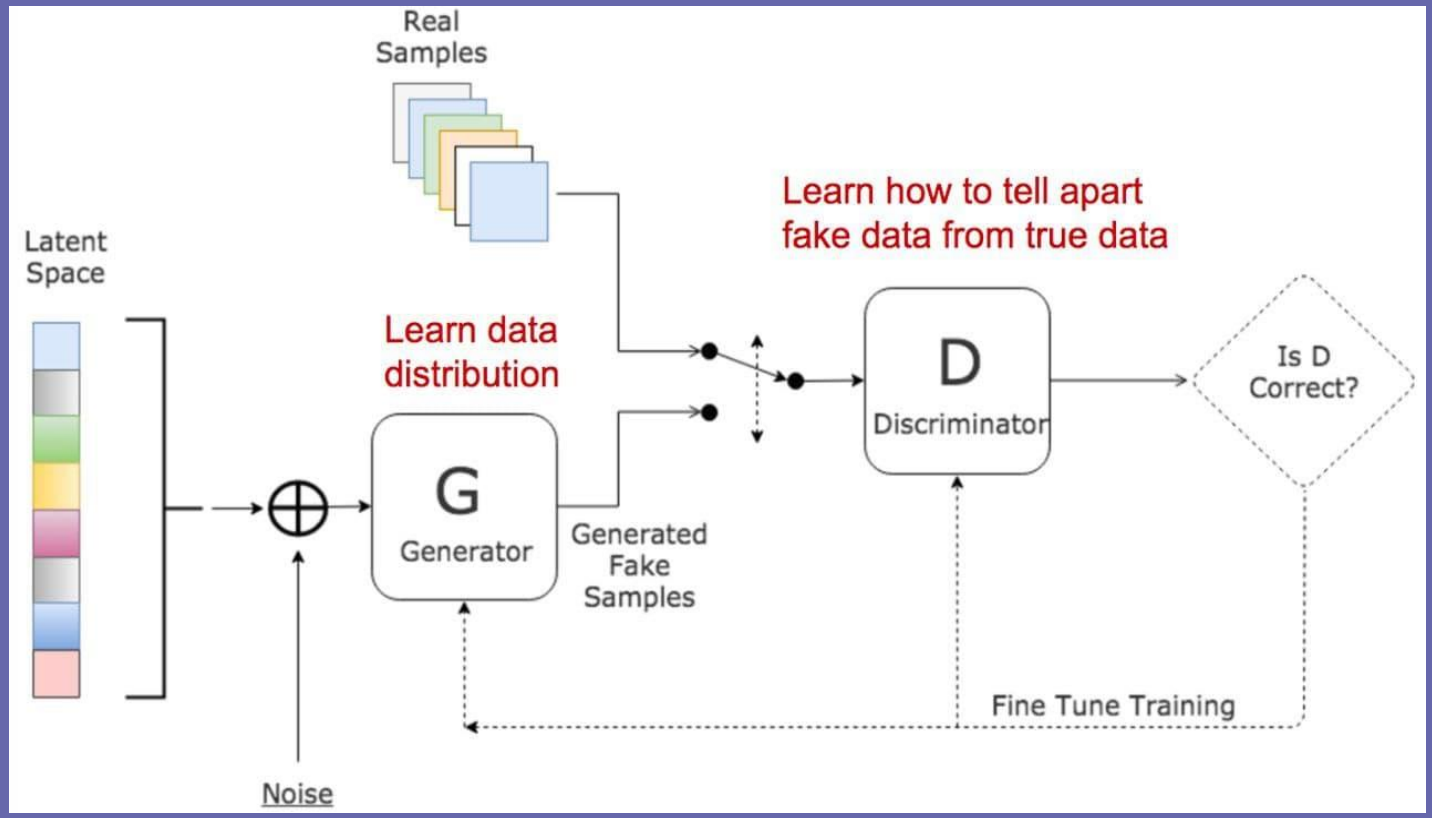
# کاربرد آموزش ۲ شبکه در یک مدل چیست ؟

این دو شبکه ، به نوعی بازی دزد و پلیس را اجرا میکنند. شبکه Generative داده های fake ولی بسیار واقعگرایانه تولید میکند در حالی که وظیفه شبکه Discriminative ، تشخیص این داده ها از داده های واقعی است. به نوعی در توضیح این شبکه ها اینگونه بیان میشود که شبکه مولد وظیفه تولید پول تقلبی را دارد درحالیکه شبکه متمایزگر وظیفه تشخیص پول های تقلبی از پول های واقعی را دارد.

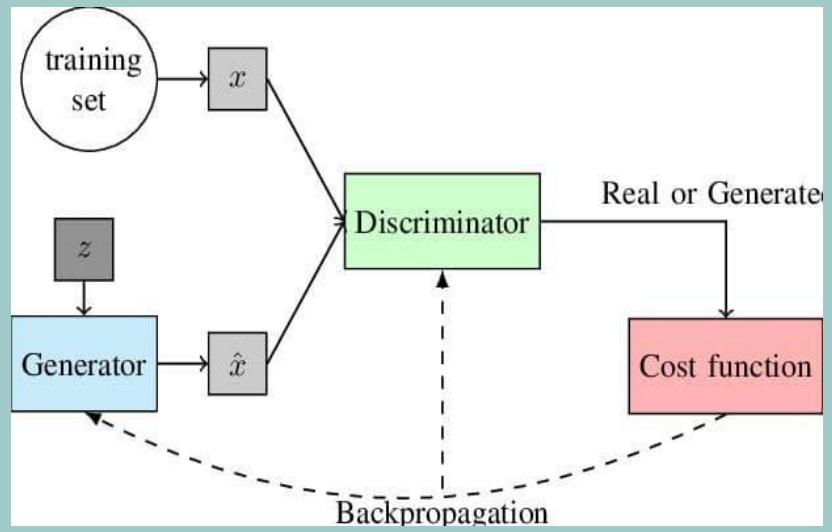


پس مدل یادگیری شبکه های مولد تخاصمی یا Generative Adversarial Networks به شرح مقابل است :

دو شبکه که اولی مولد و دومی متمایزگر است را آموزش میدهیم و سپس به کمک شبکه مولد ، داده های جدید و مطابق با واقعیت تولید میکنیم درحالیکه شبکه متمایزگر باید تفاوت این داده خیالی با داده واقعی را تشخیص بدهد.
تمرکز این آموزش بر روی بررسی تک تک این مدل ها نمیباشد بلکه هدف بررسی این دو مدل در کنار هم و به عنوان یک سیستم واحد و در ارتباط با هم میباشد. همچنین کاربرد شبکه های GAN نیز مورد بررسی قرار میگیرد.



# تعریف تابع هزینه :




جهت بررسی هر سیستم و برای شناخت بهتر هر مجموعه در ابتدا میبایست تابع هدف(هزینه) آن را تعریف نماییم. جهت تعریف تابع هدف(هزینه) نیز ابتدا میبایست به مفاهیم و درک خود از این دو سیستم بازگردیم.

هدف در اینجا بهبود کیفیت داده های تولید شده توسط مدل مولد و افزایش توانایی تشخیص داده های واقعی از غیر واقعی در مدل متمایزگر میباشد. پس به نوعی هدف افزایش دقت در مدل مولد و افزایش دقت ناشی از خروجی مدل مولد در مدل متمایزگر میباشد. اکنون با دانستن این موضوع میتوان تابع هدف را در مدل مولد و مدل متمایزگر به تفکیک زیر تعریف کرد.

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \left[ \log D\left(x^{(i)}\right) + \log\left(1 - D\left(G\left(z^{(i)}\right)\right)\right) \right].$$

رابطه تابع هزینه در شبکه متمایز گر

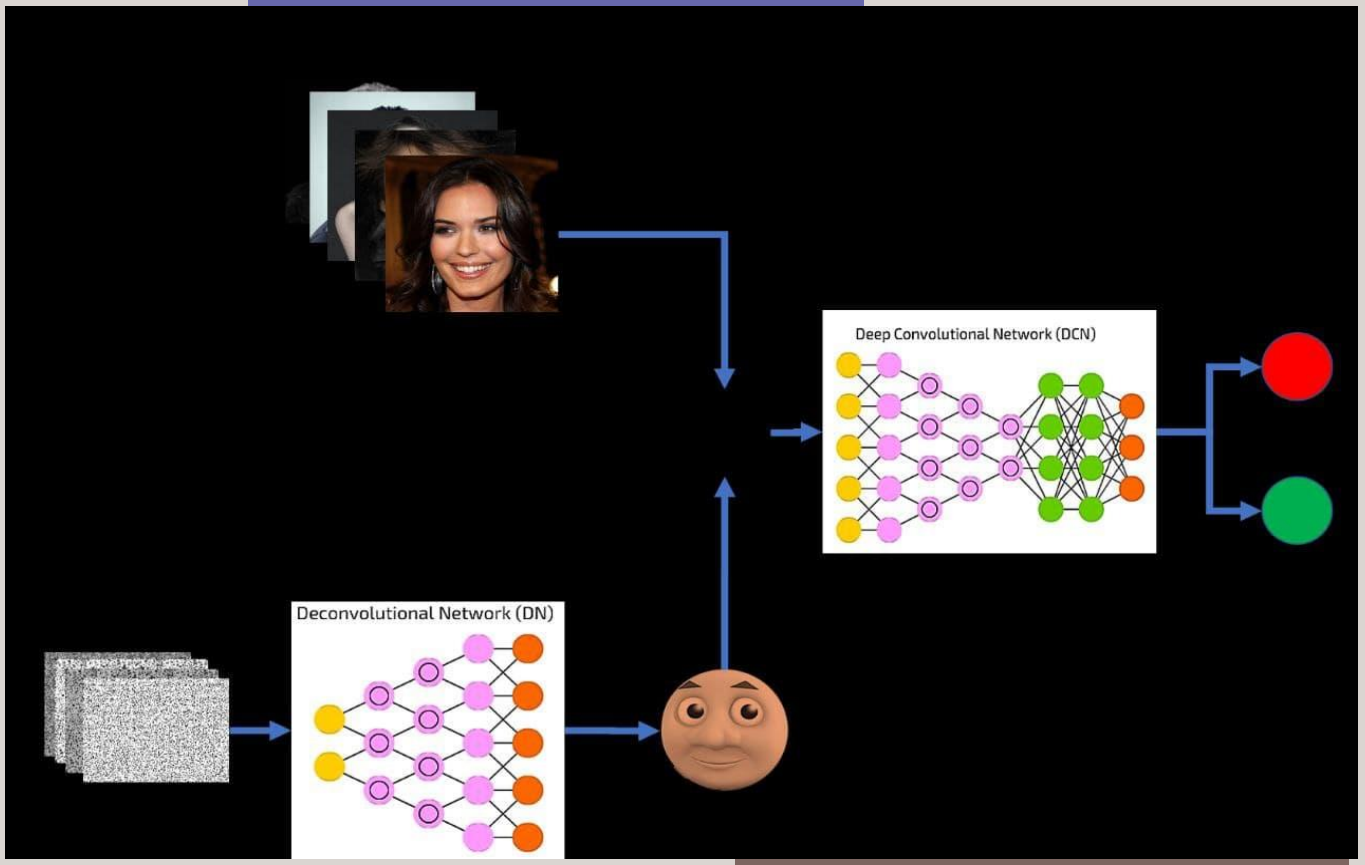
رابطه تابع هدف در شبکه مولد

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \log\left(1 - D\left(G\left(z^{(i)}\right)\right)\right).$$

که رابطه اول میبایست ماکسیمم و رابطه دوم می بایست مینیمم شود.

# شرط همگرایی مدل :

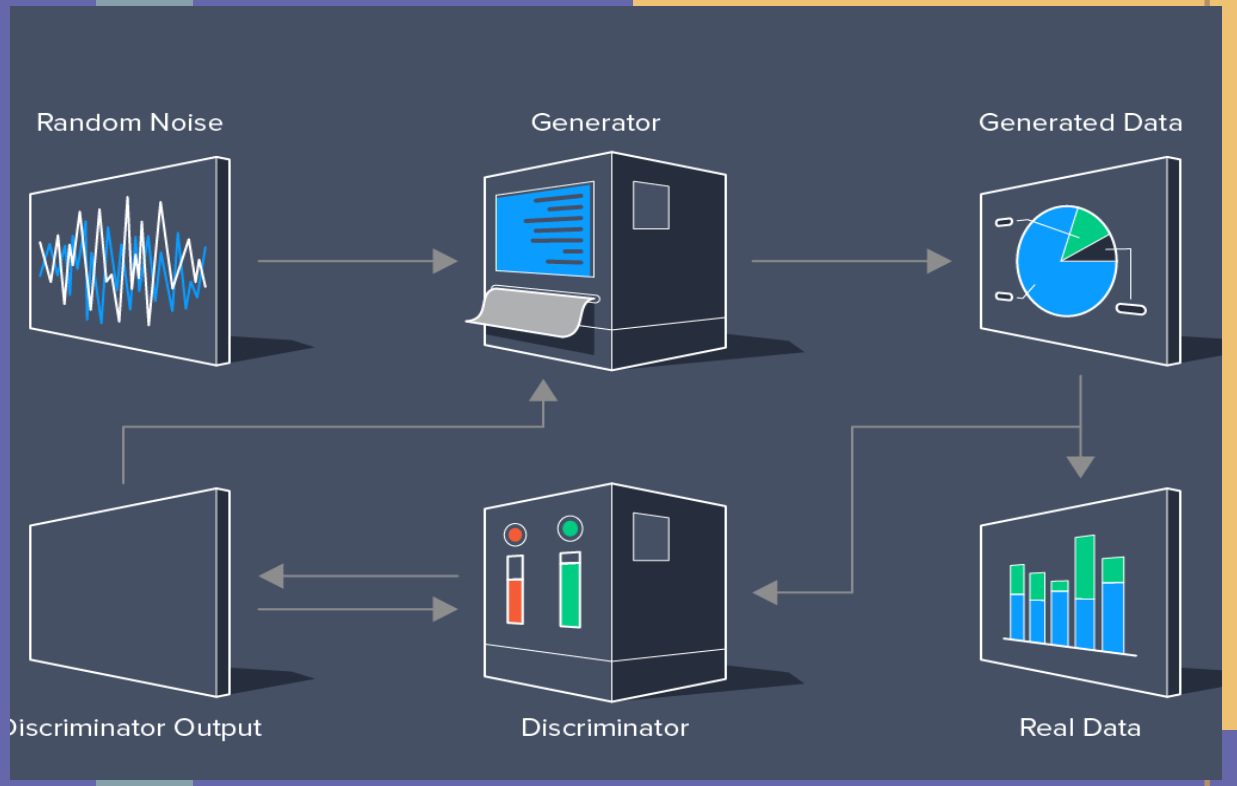
اگر ۵۰ درصد داده های فیک توسط شبکه متمایزگر تشخیص داده نشوند ، میگوییم سیستم در جهت حرکت به سمت نقطه بهینه است. سپس هرگاه سیستم درصد خاصی از داده ها مثلا ۹۰ درصد را اشتباه لیبل بزند ، میگوییم سیستم همگرا شده است.



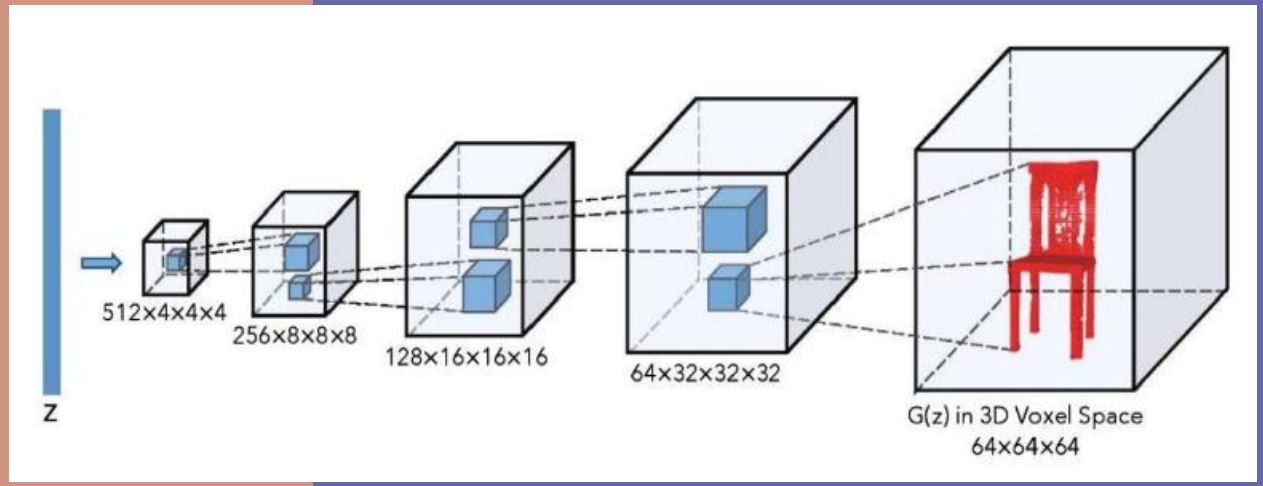
# دو شیوه رایج در تولید داده در شبکه های مولد

ابتدا ورودی رندومی را از توزیع متغیر Z که متغیر نویز است به شبکه میدهیم. (در حقیقت ورودی شبکه مولد را نویز میدهیم). سپس شبکه مولد شروع به تولید داده میکند که میتواند از یکی از روش های زیر باشد :

۱ – داده ها توسط یک شبکه عصبی پرسپترون چندلایه تولید میشوند و مثلا در کاربرد تولید عکس ، به تعداد پیکسل ها میتوان خروجی گرفت. نحوه محاسبه هر خروجی هم متناسب با توزیع داده های آموزشی است.



۲ – در روش دوم ، داده ها توسط یک شبکه عصبی کانوولوشنال افزایشی تولید میشود که شکل زیر این مفهوم را بهتر منتقل میکند :



# کاربرد های GAN

- در صنعت بازی سازی : شرکت Nvidia در سال گذشته ، تکنولوژی به اسم GauGan2 معرفی کرد که میتوان در صنعت گیمینگ تحولی ایجاد کند. در کلیپ زیر ، این کاربرد را مشاهده خواهید کرد :
در تشخیص چهره : امروزه عکس های تقلبی و غیر واقعی برای گول زدن سیستم های تشخیص چهره فراوان است. با سیستم GAN میتوان داده های فیک تولید کرد و شبکه را با آن آموزش داد که پس از آن دیگر با داده های فیک و غیرواقعی قابل دور زدن نباشد.
در پزشکی : در پزشکی از شبکه مولد میتوان جهت تولید عکس های MRI استفاده کرد.

بازیابی متون و عکس های از دست رفته :

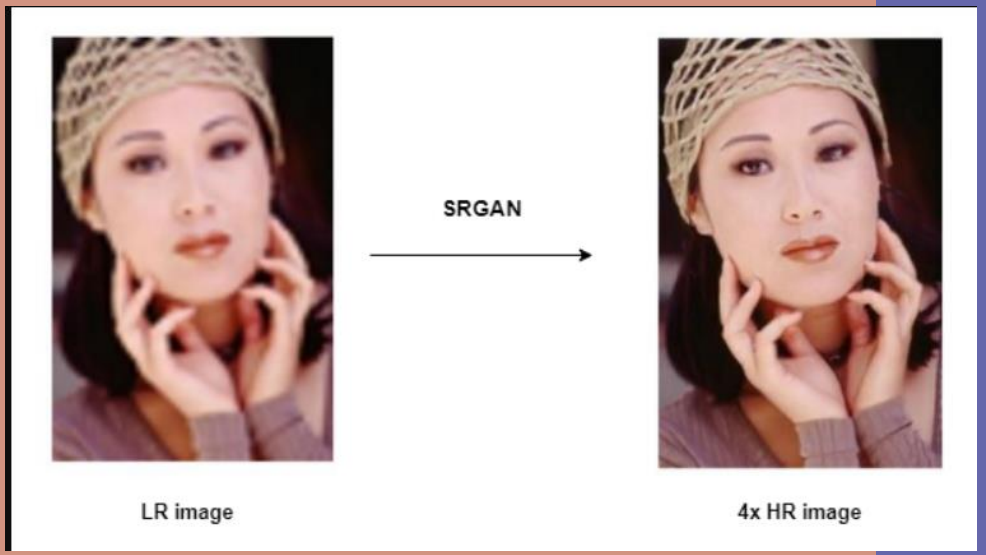
- عکس های تاریخی که از گذشته در دسترس است ، میتواند مخدوش شده باشد. با استفاده از شبکه های مولد میتوان تکه های از دست رفته را بازیابی کرد که یک نمونه آن را در زیر مشاهد میکنید :


- رنگ کردن اشیا :

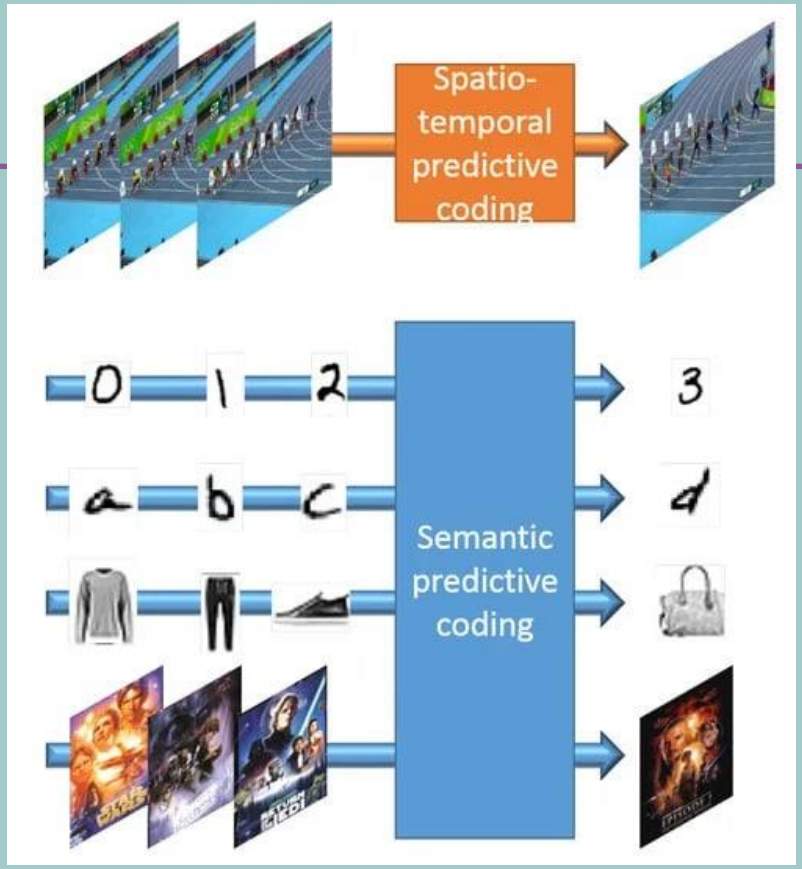
- در صنعت مدلینگ : میتوان چهره افراد جدیدی را توسط این شبکه تولید کرد و ما را از مدل های جدید بی نیاز میکند.
دیپ فیک : میتوان چهره فردی را بر روی حالات صورت فردی دیگر مونتاژ کرد.

- تبدیل تصاویر وضوح کم به تصاویر با وضوح زیاد

- SRGAN ( Super-resolution GAN) از ترکیب یک شبکه عمیق با یک شبکه دشمن برای تولید تصاویر با وضوح بالاتراستفاده می‌کند. در طول آموزش، یک تصویر با وضوح بالا ( HR : High Resolution ) به یک تصویر با وضوح پایین ( LR : Low Resolution ) با استفاده از تکنیک‌های Downsampling تبدیل می‌شود. سپس GAN Generator تصاویر LR را با تکنیک های Upsamplingبه تصاویر با وضوح فوق العاده ( SR : Super Resolution ) تبدیل می کند. پس از آن تصویر به تفکیک کننده و تمایز دهنده ( Discriminator ) داده می‌شود و آن سعی می کند بین تصویر تولید شدهٔ SRو تصویر HRتمایز قائل شود و مقدار تابع هزینهٔ استفاده شده را به معماری Generator بازمی‌گرداند. از شبکه‌ی کانولوشن لایه لایه به عنوان Discriminator استفاده می‌شود.



- پیش‌بینی و تولید ویدیو ( Video Prediction and Generation )
- درک حرکات شی و دینامیک صحنه یک مشکل اصلی در بینایی ماشین است.




برای هر دو مورد یعنی تشخیص ویدیو (به عنوان مثال، کلسیفیکیشن) و تولید ویدیو (به عنوان مثال، پیش بینی آینده)، مدلی از نحوه ی تبدیل صحنه ها مورد نیاز است.

با این حال، ایجاد یک مدل دینامیکی چالش برانگیز است چرا که تعداد زیادی راه وجود دارد که اشیا و صحنه ها می توانند تغییر کنند.

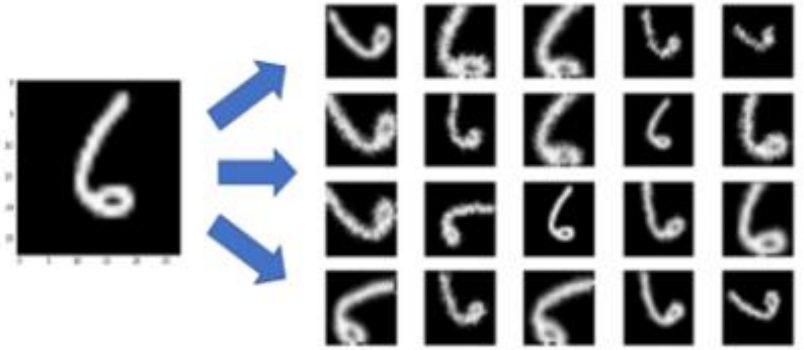
به عنوان مثال یک شبکه مولد تخاصمی با معماری کانولوشنال مکانی-زمانی می تواند پیش‌زمینه صحنه را از پس‌زمینه جدا کند.آزمایش‌ها نشان می‌دهد که این مدل می‌تواند ویدیوهای کوچکی را تا یک ثانیه با فریم کامل تولید کند و کاربرد آن را در پیش‌بینی آینده ی قابل قبول تصاویر ثابت (static images)، نشان می‌دهد.

# مشکلات رایج GAN :

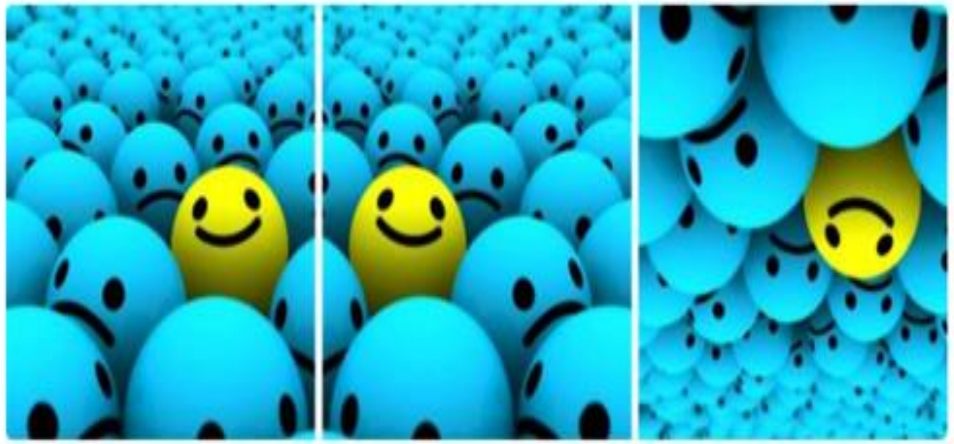
۱ – مشکل اول : شبکه های مولد جهت آموزش نیاز به تعداد زیادی داده دارند. مثلا در کاربرد تصویر ، جهت آموزش نیاز به ده ها میلیون تصویر با کیفیت دارند. طبیعتا پیدا کردن این تعداد تصویر برای آموزش شبکه ، کار بسیار دشواریست. پس این اولین چالش در آموزش شبکه های مولد میباشد.

راه حل : راه حلی به نام داده افزایی یا Data Augmentation جواب مسئله است. در این تکنیک ، یک داده را به شکل های مختلفی با تغییرات اندک تکرار میکنند(مثل چرخش و ...) و با این روش ، داده افزایی میکنند.
داده افزایی روش های مختلفی دارد که در ادامه به چند مورد از آن اشاره میکنیم :

1- چرخش



2- Flip

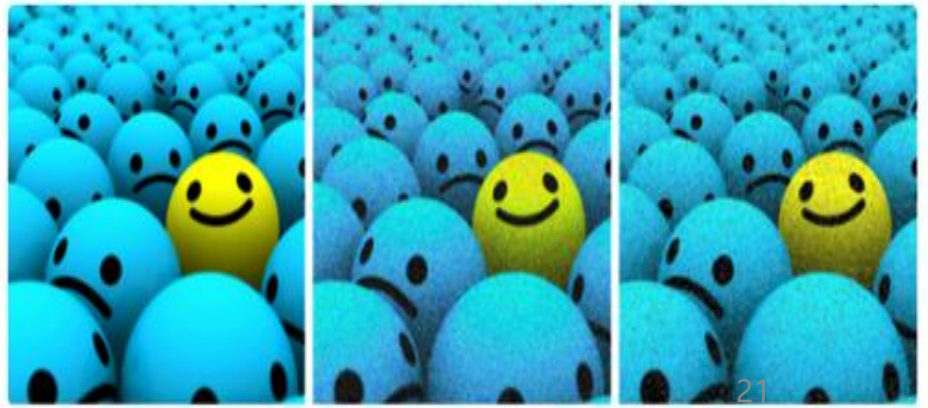


5- تغییر رنگ عکس

6- تغییر روشنایی عکس

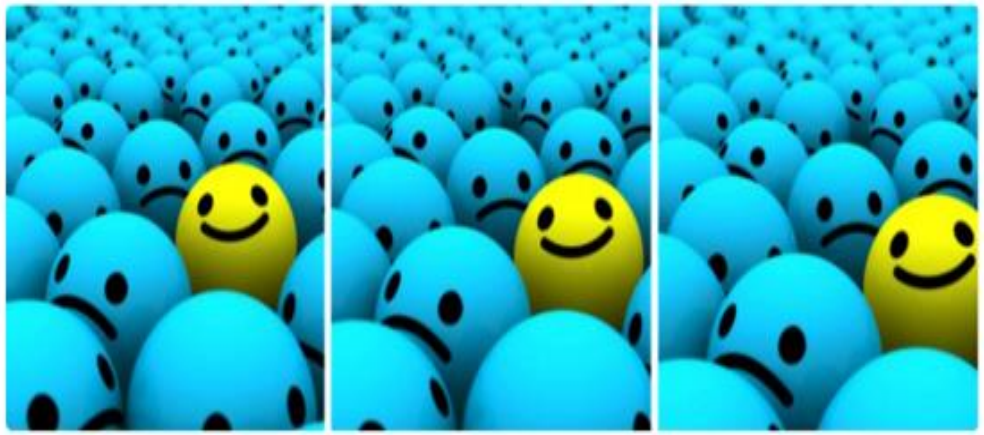
7- .....

3- اضافه کردن نویز گاوسی یا نمک فلفل



From the left, we have the original image, image with added Gaussian noise, image with added salt and pepper noise

4- زوم کردن



این روش ها و ده ها روش دیگر همگی در کنار هم میتوانند باعث داده افزایی شوند و میتوان گفت بجای میلیون ها عکس ، میتوان با یک دهم تعداد عکس در مرحله قبل ، شبکه را آموزش داد.
قابل ذکر است که این تکنیک نه تنها در تصویر ، بلکه در هر نوع داده دیگری مثل متن نیز قابل استفاده است.
نکته جالب تر آنکه شبکه های مولد ، خود برای داده افزایی نیز استفاده میشوند.
۲- دومین مشکل رایج در آموزش ، از راه حل مشکل اول ناشی میشود. اگر تعداد زیادی داده غیرسالم به شبکه بدهیم ، چون شبکه تبدیل هایی مثل چرخش و ... را یاد میگیرد ، منجر به تولید داده های ناسالم میشود .
راه حل : طبق تجربه و تحقیقات ، اگر تعداد داده های سالم در برابر کل داده ها ، بیشتر از ۲۰ درصد باشد، شبکه تا حد بسیار قابل قبولی ، داده های سالم تولید میکند.

# با تشکر از توجه شما