

دانشگاه صنعتي امیرکبیر  
(پلی تکنیک تهران)

دانشكده مهندسی کامپیوتر

سمینار درس مبانی یادگیری آماری کارشناسی‌ارشد

مدل‌های احتمالاتی پخشی کاهش نویز

نگارش

حمید رضایی

استاد درس

دکتر مریم امیرمزلقانی

ماه و سال

تیر 1403

صفحه فرم ارزیابی و تصویب پایان نامه- فرم تأیید اعضاء كميته دفاع

در این صفحه(هر سه مقطع تحصيلي) بايد تصویر فرم ارزيابي یا تأیید و تصویب پایان­نامه/رساله موسوم به فرم کمیته دفاع براي مقاطع­ کارشناسی­ارشد و دكتری و تصویر فرم تصويب براي مقطع كارشناسي، موجود در پرونده آموزشی را قرار دهند.

اينجانب حمید رضایی متعهد مي‌شوم كه مطالب مندرج در اين پايان نامه حاصل كار پژوهشي اينجانب تحت نظارت و راهنمايي اساتيد دانشگاه صنعتي اميركبير بوده و به دستاوردهاي ديگران كه در اين پژوهش از آنها استفاده شده است مطابق مقررات و روال متعارف ارجاع و در فهرست منابع و مآخذ ذكر گرديده است. اين پایان نامه قبلاً براي احراز هيچ مدرك هم‌سطح يا بالاتر ارائه نگرديده است.

در صورت اثبات تخلف در هر زمان، مدرك تحصيلي صادر شده توسط دانشگاه از درجه اعتبار ساقط بوده و دانشگاه حق پيگيري قانوني خواهد داشت.

كليه نتايج و حقوق حاصل از اين پایان نامه متعلق به دانشگاه صنعتي اميركبير مي‌باشد. هرگونه استفاده از نتايج علمي و عملي، واگذاري اطلاعات به ديگران يا چاپ و تكثير، نسخه‌برداري، ترجمه و اقتباس از اين پایان نامه بدون موافقت كتبي دانشگاه صنعتي اميركبير ممنوع است.   
نقل مطالب با ذكر مآخذ بلامانع است.

در صفحه تعهدنامه اصالت اثر، در قسمت بالا سمت چپ، تاریخ دفاع خود را جایگزین تاریخ نوشته شده کنید.

همچنین در صفحه تعهدنامه اصالت اثر، در خط اول، نام و نام خانوادگی خود را به صورت کامل با نام و نام خانوادگی نمونه، جایگزین کنید. در انتهای متن تعهد، در قسمت امضا نیز باید نام و نام خانوادگی کامل خود را وارد نماید.

حمید رضایی

امضا

**در صورت تمایل این صفحات** (به صورت صفحات مجزا) **نیز اضافه شود: (اختیاری)**

* **صفحه تقديم**

نويسنده پايان‏نامه، درصورت تمايل مي‏تواند برای سپاسگزاری پايان‏نامه خود را به شخص يا اشخاص و يا ارگان خاصی تقدیم نماید.

* **صفحه تقدير و تشكر**

نويسنده پايان‏نامه مي‏تواند مراتب امتنان خود را نسبت به استادراهنما و استادمشاور و یا ديگر افرادي كه طي انجام پايان‏نامه به نحوي او را یاری و یا با او همكاري نموده‏اند ابراز دارد.

# چكيده

در این گزارش سعی شده است یکی از جدیدترین روش‌های ساخت تصاویر مصنوعی برپایه مدل‌های آمار و احتمالاتی مورد برسی قرار گیرد، مدل‌های پخشی احتمالاتی[[1]](#footnote-1) یکی از به‌روزترین مدل‌های مولد تصاویر مصنوعی هستند. رفتار این مدل‌های متغیر پنهان با نویز در اصل مشابه با رفتار سیستم‌ها در ترمودینامیک غیرتعادلی است. این مدل‌ها بهترین نتایج خود را با آموزش بر روی یک کران واریانسی وزنی به دست آورده است که این کران از یافتن یک ارتباط جدید بین مدل‌های احتمالاتی پخشی و تطبیق امتیاز کاهش نویز به کمک دینامیک لانژوین [[2]](#footnote-2)به دست آمده است.

واژه‌های کلیدی:

مدل‌های احتمالاتی، فرآیندهای مارکوف، مدل‌های مولد

|  |  |
| --- | --- |
| فهرست مطالب | صفحه |

[چكيده ‌أ](#_Toc171639056)

[فصل اول مقدمه مقدمه 1](#_Toc171639057)

[فصل دوم نگاهی عمیق‌تر به دیفیوژن مدل‌ها 5](#_Toc171639058)

[پیش‌زمینه 6](#_Toc171639059)

[2-1- فرآیند پخش رو به جلو 6](#_Toc171639060)

[2-2- معکوس فرآیند پخشی 8](#_Toc171639061)

[2-3- تابع هزینه 12](#_Toc171639062)

[2-3-1- ساده‌سازی 13](#_Toc171639063)

[2-4- مقیاس داده‌های ورودی 14](#_Toc171639064)

[فصل سوم آزمایش 16](#_Toc171639065)

[آزمایش 17](#_Toc171639066)

[3-1- روش ارزیابی 18](#_Toc171639067)

[فصل چهارم معماری مدل U-Net 19](#_Toc171639068)

[فصل پنجم جمع‌بندي و نتيجه‌گيري جمع‌بندي و نتيجه‌گيري 22](#_Toc171639069)

[منابع و مراجع 24](#_Toc171639070)

[پيوست‌ها 25](#_Toc171639071)

[Abstract 26](#_Toc171639072)

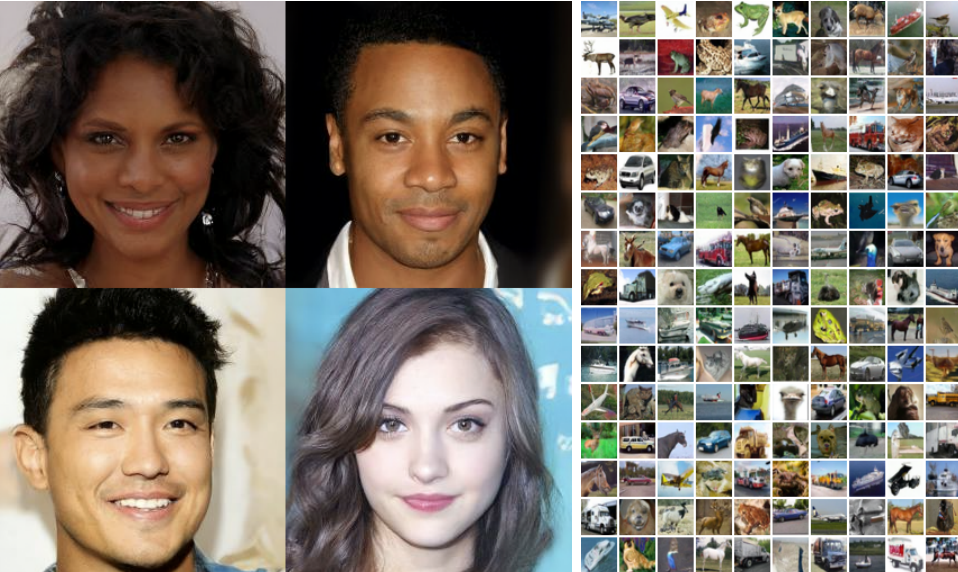
# فصل اول مقدمه مقدمه

تاکنون، مدل‌های مولد متنوعی همچون شبکه‌های مولد تخاصمی (GAN)، مدل‌های خودرمزگذار متغیر (VAE) و مدل‌های مبتنی بر جریان (Flow-based) معرفی شده‌اند. با به‌کارگیری روش‌هایی همچون مدل‌سازی بر اساس انرژی و تطبیق امتیاز، پیشرفت‌های چشمگیری نیز در تولید تصاویر مصنوعی با کیفیت بالا حاصل شده است. با این حال، هر یک از این مدل‌ها محدودیت‌های خاص خود را دارند. به‌عنوان مثال، شبکه‌های مولد تخاصمی به دلیل ذات تخاصمی خود، دارای یادگیری غیرپایدار هستند و همچنین تصاویر تولید شده توسط این مدل‌ها دارای تنوع پایینی است.

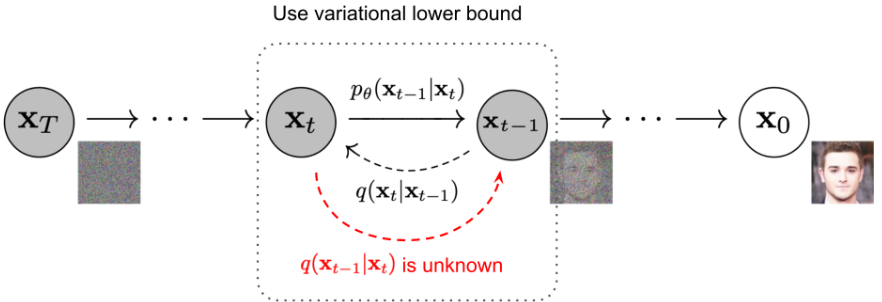
در این راستا، مدل‌های پخشی دسته دیگری از مدل‌های مولد هستند که از ترمودینامیک غیرتعادلی الهام گرفته‌اند. این مدل‌ها یک زنجیره مارکوف پارامتری هستند که با استفاده از استنتاج واریانس آموزش داده می‌شوند تا پس از یک زمان محدود، نمونه‌هایی مشابه با داده‌ها تولید کنند. مدل‌های پخشی به‌طور معمول شامل دو فرایند اصلی هستند: فرایند انتشار (diffusion process) که داده‌ها را به نویز تصادفی تبدیل می‌کند، و فرایند معکوس (reverse process) که نویز را به داده‌های واقعی بازمی‌گرداند.

یکی از ویژگی‌های برجسته مدل‌های پخشی، پایداری یادگیری و توانایی بالای آن‌ها در تولید نمونه‌های متنوع است. این مدل‌ها با تعریف فرایند انتشار به‌صورت تدریجی و پارامتری کردن این فرایند، می‌توانند تصاویر با کیفیت و تنوع بالا تولید کنند. علاوه بر این، مدل‌های پخشی به دلیل ساختار خود، قابلیت تفسیرپذیری بالایی دارند و می‌توانند به‌خوبی درک شوند.

در نتیجه، مدل‌های پخشی به‌عنوان یک رویکرد جدید و نوآورانه در تولید تصاویر مصنوعی، توانسته‌اند جایگاه ویژه‌ای در بین مدل‌های مولد پیدا کنند و به‌طور چشمگیری محدودیت‌های موجود در مدل‌های پیشین را کاهش دهند. با توجه به این ویژگی‌ها، این مدل‌ها می‌توانند به‌طور گسترده در کاربردهای مختلفی همچون تولید تصاویر با کیفیت بالا، بازسازی تصاویر، و ایجاد داده‌های مصنوعی برای آموزش مدل‌های یادگیری عمیق مورد استفاده قرار گیرند.



شکل 1 نمونه‌های تولید شده توسط مدل دیفیوژن روی دیتاست CelebA-HQ 256 × 256 (تصاویر سمت چپ) و دیتاست بدون شرط CIFAR10 (تصاویر سمت راست) [1]



شکل 2 زنجیره مارکوف [[3]](#footnote-3)پخشی با کمی تغییر [1]

فرآیند یادگیری در اصل بدین گونه است که در یک زنجیره مارکوف ابتدا به تصویر اولیه بعد از هر مرحله مقداری نویز اضافه می‌شود تا زمانی که پس از تعداد گام مناسب در نهایت تصویر کاملا از بین رفته و نویز گوسی با میانگین صفر و واریناس یک ایجاد می‌شود. سپس در فرآیند برگشت سعی می‌شود تا احتمال انتقال از تصویر نویزی به تصویر بدون نویز را با استفاده از شبکه‌ عصبی تخمین زد. به بیان دیگر، شبکه عصبی تلاش می‌کند تا فرآیند معکوس را مدل‌سازی کند و نویز را به یک تصویر واقعی و با کیفیت بازگرداند.

در رابطه با زنجیره مارکوفی که آموزش داده می‌شود نکات مهمی که باید درنظر گرفت به ترتیب عبارت اند از[2]:

1- متغیرهای پنهان در فرآیند مارکوف ابعاد بزرگی دارند: ابعاد این متغیرها برابر با ابعاد تصاویر اولیه است. به عبارت دیگر، هر نقطه در تصویر اولیه یک متغیر پنهان در زنجیره مارکوف محسوب می‌شود.

2- فرض می‌شود که توزیع انتقال نرمال است (در واقع داده در لحظه t-1‌ دارای توزیع نرمال است) و دلیل این فرض این است که گام‌های کوچک کوچکی برداشته شده اند و این گام‌های کوچک تضمین می‌کنند که داده در یک لحظه قبل‌تر نیز گوسی بوده است. به این دلیل است که به تصویر مرحله به مرحله نویز اضافه می‌شود و یک دفعه تصویر را در یک مرحله نویزی نمی‌کنیم.

# فصل دوم نگاهی عمیق‌تر به دیفیوژن مدل‌ها

# پیش‌زمینه

در ابتدا برای فهم بهتر اینکه دیفیوژن مدل‌ها چگونه کار می‌کنند و فرآیند یادگیری آنها چگونه است، باید چند مورد مورد از مفاهیم آشنا شد و درک خوبی از تئوری و ریاضی پشت هر یک داشت.

اولین موردی که باید به آن توجه داشت این است که در حوزه آماری فرآیند پخشی (Diffusion) به چه معناست، یک فرآیند پخشی یک فرآیند مارکوف تصادفی است که دارای مسیر نمونه پیوسته می‌باشد که در طی این مسیر از یک توزیع پیچیده همانند توزیع دیتاست به یک توزیع ساده‌تر مانند توزیع نرمال می‌رسیم.

## فرآیند پخش رو به جلو[[4]](#footnote-4)

فرض کنید یک داده از دیتاست واقعی نمونه برداری شده است، x0 ∼ q(x0)، برای پیاده سازی فرآیند پخش روی آن در T مرحله مقدار کمی نویز به داده اولیه اضافه می‌کنیم، که در نتیجه یک دنباله از نمونه‌های نویزی، ، ایجاد می‌شود. در این فرآیند اندازه گام‌ها توسط واریانس‌های β1  **تا** βT تعیین می‌شوند.

با توجه به فرمول زیر ثابت می‌شود با گذشت زمان به اندازه کافی بطوری که T به قدر کافی بزرگ باشد داده اولیه به تدریج ویژگی‌های قابل تشخیص خود را از دست می‌دهد و درنهایت توزیع برابر می‌شود با توزیع نرمال.

نکته‌ای که باید درنظر گرفت این است که مقدار β ثابت نیست و در فرآیند پخش مقدار آن با افزایش t افزایش می‌یابد. در فرآیند روبه جلو مقدار واریانس βt را می‌توان به صورت پارامتر یادگرفت و یا اینکه بصورت هایپرپارامتر[[5]](#footnote-5) یک مقدار ثابت به آن نسبت داد. که در مقاله [1] مقدار واریانس ثابت در نظر گرفته شده است و مقدار میانگین برای هر حالت یادگرفته می‌شود.

## معکوس فرآیند پخشی[[6]](#footnote-6)

۹اگر بتوانیم فرآیند قبلی که بحث شد را معکوس کنیم و از نمونه برداری کنیم، می‌توانیم نمونه‌ی واقعی را از ورودی نویز گوسی بازسازی کنیم. اما نکته مهمی که باید درنظر گرفت این است که، ما نمی‌توانیم به راحتی احتمالات شرطی را تخمین بزنیم زیرا نیاز به استفاده از کل مجموعه داده داریم. به بیان دیگر باید از کل دنیا شناخت داشته باشیم و بدونیم دادگان اصلی به چه شکل بوده اند. متاسفانه این امر بسیار دشوار و تا حدی غیر ممکن است. به همین دلیل در فرآیند معکوس مدل را که تقریبی از احتمالات شرطی بالا است را آموزش می‌دهیم.

باید توجه داشت که احتمال شرطی معکوس قابل محاسبه است زمانی که شرط آن به x0  وابسته باشد:

با استفاده از قانون بیز[[7]](#footnote-7) داریم:

که در رابطه بالا یک تابع است که به  بستگی ندارد. پس با توجه به تابع توزیع گوسی، میانگین و واریانس را می‌توان به صورت زیر نوشت.

با توجه به و داریم:

حال اگر را جایگذاری کنیم داریم:

حال که فرآیند برگشتی را بصورت پارامتری با فرم بسته نوشتیم نوبت این است که تابع هزینه را نیز بدست آوریم. در مدل‌های دیفیوژن نیز همانند VAE از حد پایین واریانس برای بهینه کردن منفی log-likelihood استفاده می‌شود.

پس داریم:

حال برای اینکه محاسبه هر جز از معادله فوق ساده‌تر شود می‌توان تابع هدف را طوری باز نویسی کرد که برابر باشد با مجموع چند جز KL-divergence:

با توجه به قانون بیز داریم:

پس درنهایت داریم به فرمول ساده شده زیر می‌رسیم.

که:

از آنجایی که تمامی KL-divergenceهای فرمول بالا دو توزیع گوسی را مقایسه می‌کنند پس می‌توان برای آنها فرم بسته نوشت .[3]

از آنجایی که هیچ پارامتر قابل یادگیری‌ای ندارد و یک نویز گوسی است، پس ثابت است و می‌توان آن را درنظر نگرفت.

## تابع هزینه

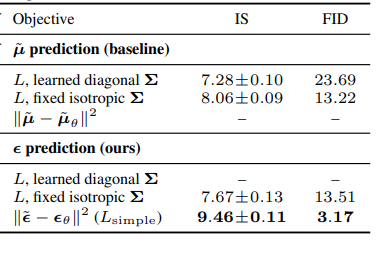
همانطور پیش‌تر بحث شد برای تقریب زدن توزیع احتمال شرطی‌ای که در معکوس فرآیند پخش وجود دارد، ، نیاز به آموزش یک شبکه عصبی است. بدین منظور باید را با توسط شبکه عصبی تخمین زد. توجه داشته باشید که مقدار واریانس را آموزش نمی‌دهیم و به عنوان هاپرپارامتر به خورد شبکه می‌دهیم.

نکته مهمی که باید درنظر گرفت این است که از آنجایی که به عنوان ورودی در زمان آموزش موجود است، می‌توان با کمی بازی با ریاضی نویز گوسی را مدل کنیم تا را تخمین بزنیم.

با توجه به رابطه بالا داریم:

مقدار هزینه را باید طوری باشد که اختلاف و کم‌ترین مقدار ممکن بشود.

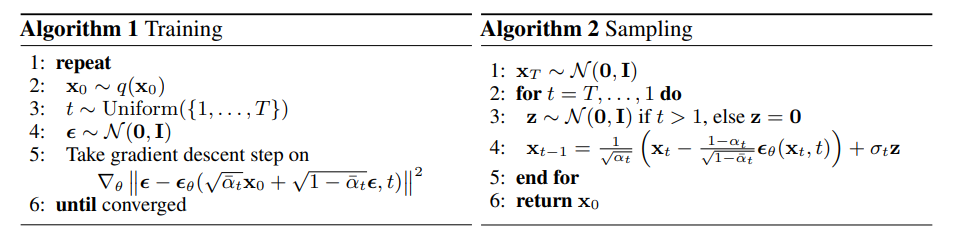
### ساده‌سازی



شکل 3: نتایج به‌دست آمده در آزمایشات [1]

در مقاله [1] بصورت تجربی به این نتیجه رسیده شده که درصورت ساده‌سازی و عدم استفاده از وزن‌ها نتایج بهتری بدست می‌آیند. پس درنهایت داریم:

که مقدار ثابت است و به θ بستگی ندارد.



شکل 4: الگوریتم آموزش و نمونه برداری[1]

## مقیاس داده‌های ورودی

با فرض اینکه داده‌های ورودی اعدادی بین 0 تا 255 باشند، آنها را بصورت خطی بین 1- تا 1 مقایس می‌کنند تا ورودی‌ها در شبکه عصبی به صورت یکنواختی مقیاس‌بندی شوند. این کار برای این انجام می‌شود که فرآیند معکوس (رمزگشایی) که در مدل‌های مولد استفاده می‌شود، به خوبی عمل کند.

نکته دیگری که باید درنظر گرفت این است که برای به دست آوردن احتمال لگاریتمی گسسته، آخرین عبارت فرآیند معکوس را به صورت یک رمزگشای گسسته مستقل مشتق شده از گوسی تنظیم می‌شود:

که ابعاد داده است و نیز نشان دهنده یک مختصات از کل ابعاد است.

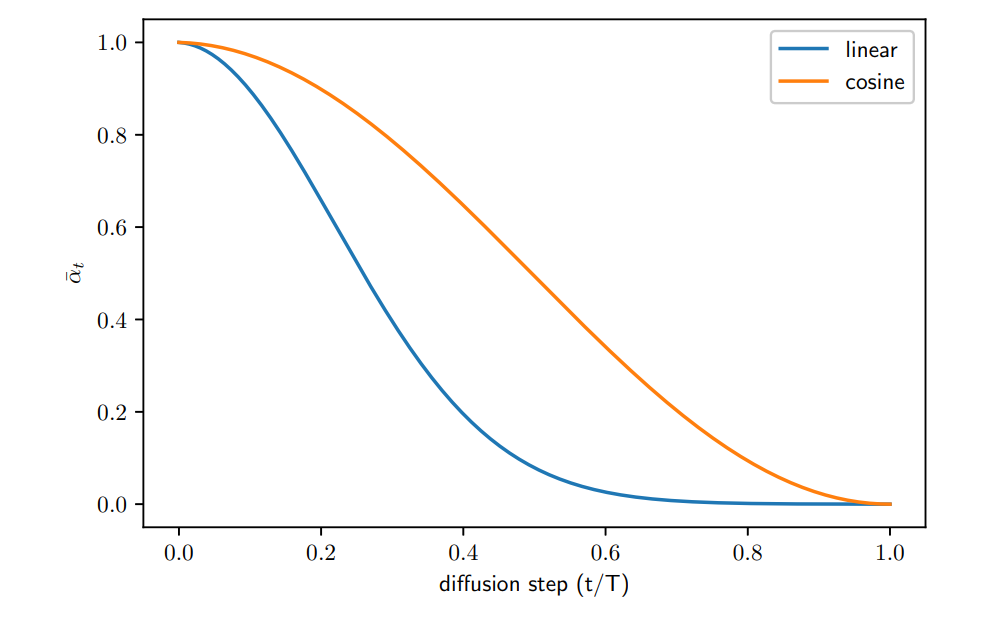
# فصل سوم آزمایش

# آزمایش

در تمام آزمایشات انجام شده در مقاله [1] مقدار فرض شده است. و همچنین مقدار واریانس مرحله رو به جلو دنباله‌ای خطی افزایشی از اعداد ثابت است، بطوری که و .

این مقدار در مقایسه با داده‌های نرمال شده که در بازه هستند. و دلیل آن این است که در هر دو فرآیند رو به جلو و معکوس سعی بر این بوده است تا نسبت سیگنال به نویز در کم‌ترین مقدار ممکن را داشته باشد.

در مقاله [2] که سعی دارد پیشرفتی برای مدل ارائه شده توسط مقاله [1] ارائه دهد یکی از کارهایی که صورت گرفته است این است که به جای اینکه مقدار واریانس را بصورت خطی افزایش دهد بصورت کوسینوسی این کار را انجام می‌دهد.



شکل 5: مقایسه زمان‌بندی خطی و بر اساس کسینوس واریانس در طول آموزش[2]

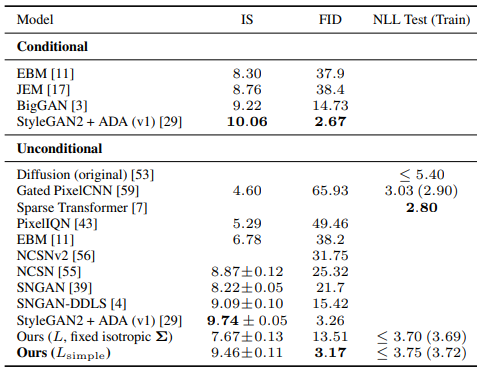
## روش ارزیابی

در مقاله [1] از سه معیار و و استفاده شده است. که به ترتیب برای هر یک توضیح مختصر داده خواهد شد.

1- Inception score: IS از دو جنبه مدل را ارزیابی می‌کند: اول، انتخاب یک مجموعه از تصاویر تولید شده، و دوم، ارزیابی تنوع این تصاویر. برای این کار، از یک شبکه آموزش دیده برای دسته‌بندی تصاویر و محاسبه معیارهای شباهت بین دسته‌های مختلف استفاده می‌شود.

2- FID score: به عنوان یک معیار دیگر کیفیت تصاویر تولید شده توسط یک شبکه مولد به کار می‌رود. این معیار تفاوت دو توزیع احتمال را بین تصاویر واقعی و تصاویر تولید شده از طریق محاسبه فاصله فراخت بین توزیعات نشان می‌دهد. هرچه FID کمتر باشد، بهترین کیفیت تولید تصاویر را نشان می‌دهد.

3- negative log likelihood: این معیار میزان شباهت میان توزیع احتمال واقعی داده‌ها و توزیع احتمالی که توسط شبکه مولد پیش‌بینی می‌شود را نمایش می‌دهد. اگر مدل مولد موفق به تولید داده‌هایی با توزیع احتمال مشابه به داده‌های واقعی باشد، مقدار NLL کمتر خواهد بود.

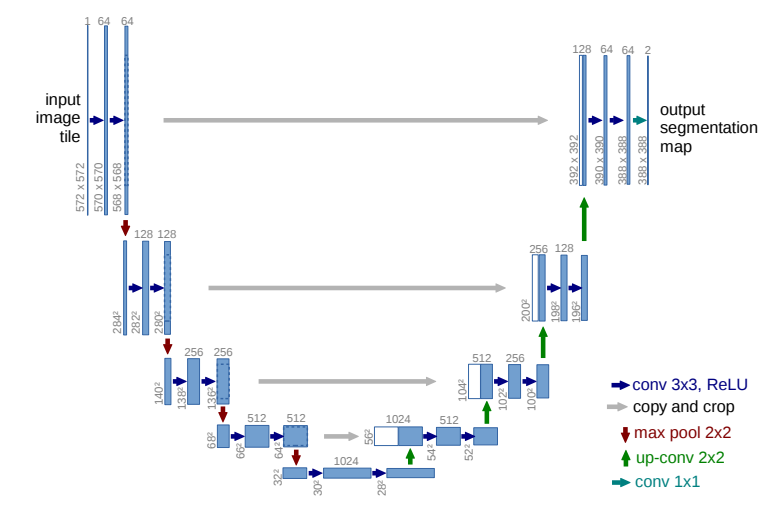


شکل 6: نتایج مدل مقاله [1] روی دیتاست CIFAR10

# فصل چهارم معماری مدل U-Net

همانطور که پیش‌تر گفته شد معاری مدل مورد استفاده در مقاله [1] U-Net است. در این معماری مسیر کاهشی همانند یک شبکه پیچشی معمولی است که از آن به منظور نمونه کاهی استفاده می‌شود و در هر گام تعداد کانال‌ها دو برابر می‌شود. از طرفی مسیر افزایشی از شبکه‌های پیچشی برای نمونه افزایی استفاده می‌شود و در هر گام تعداد کانال‌ها نصف می‌شود تا درنهایت تصویر قطعه‌بندی شده تصویر اصلی ایجاد شود.

مهم‌ترین نوآوری در این معماری در میسر افزایشی است رخ داده است، بگونه‌ای که در هر مرحله خروجی‌های گام هم ردیف از مسیر کاهشی به همراه خروجی مرحله قبلی به عنوان ورودی به گام بالاتر داده می‌شوند. بطور کلی در این مسیر خروجی‌های گام‌های پایین‌تر حاوی ویژگی‌های استخراج شده از تصویر هستند و خروجی‌های گام‌ هم‌ردیف شامل اطلاعات مکانی تصویر هستند. دلیل استفاده از خروجی گام هم ردیف این است که میسر افزایشی بتواند با کمک اطلاعات مکانی تصویر،‌ تصویر بزرگ‌تر را تشکیل بدهد.



شکل 7: معماری U-net [4]

# فصل پنجم جمع‌بندي و نتيجه‌گيري جمع‌بندي و نتيجه‌گيري

در پايان در مقاله [1] با استفاده از مدل‌های دیفیوژن، نمونه‌های تصویری با کیفیت بسیار بالا تولید شد. همچنین، ارتباطات بین مدل‌های دیفیوژن و استنتاج واریانسی برای آموزش زنجیره‌های مارکوف، تطابق امتیاز نویززدایی و دینامیک لانژوین بازپختی[[8]](#footnote-8) (و به تبع آن مدل‌های مبتنی بر انرژی)، مدل‌های خودرگرسیو[[9]](#footnote-9) و فشرده‌سازی تدریجی با افت کیفیت [[10]](#footnote-10)پیدا شده است.

مدل‌های انتشار نه تنها تصاویر با کیفیتی تولید می‌کنند، بلکه می‌توانند با روش‌ها و مدل‌های مختلف در یادگیری ماشین و تولید داده‌ها همکاری کنند. از آنجایی که این مدل‌ها به خوبی می‌توانند داده‌های تصویری را درک کنند و ساختارهای پیچیده‌ای را یاد بگیرند، محققان مقاله [1] امیدوارند که بتوان از دیفیوژن مدل‌ها در انواع دیگر داده‌ها مانند صدا، متن و... استفاده کرد.

به طور خلاصه، دیفیوژن مدل‌ها پتانسیل زیادی دارند و امید است از این پتانسیل در زمینه‌های مختلف بهره برده شود.

**منابع و مراجع**

|  |  |
| --- | --- |
|  | [1] Ho, J., Jain, A., & Abbeel, P. (2020). Denoising Diffusion Probabilistic Models. ArXiv. /abs/2006.11239 |
|  | [2] Nichol, A., & Dhariwal, P. (2021). Improved Denoising Diffusio Probabilistic Models. ArXiv. /abs/2102.09672  [3]<https://en.wikipedia.org/wiki/Kullback%E2%80%93Leibler_divergence#Multivariate_normal_distributions>  [4] Olaf Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” arXiv (Cornell University), May 2015, doi: https://doi.org/10.48550/arxiv.1505.04597. |

# پيوست‌ها

یک نمونه کد پیاده سازی دیفیوژن مدل‌ها در لینک زیر آمده است.

https://colab.research.google.com/drive/1sjy9odlSSy0RBVgMTgP7s99NXsqglsUL?usp=sharing

# Abstract



Amirkabir University of Technology  
(Tehran Polytechnic)

… Department …

MSc or PhD Thesis

Title of Thesis

By

Name

Supervisor

Dr.

Advisor

Dr.

Month & Year

1. diffusion probabilistic models [↑](#footnote-ref-1)
2. Langevin [↑](#footnote-ref-2)
3. Markov chain [↑](#footnote-ref-3)
4. Forward diffusion process [↑](#footnote-ref-4)
5. Hyperparameter [↑](#footnote-ref-5)
6. Reverse process [↑](#footnote-ref-6)
7. Bayes’ rule [↑](#footnote-ref-7)
8. annealed Langevin dynamics [↑](#footnote-ref-8)
9. autoregressive [↑](#footnote-ref-9)
10. progressive lossy compression [↑](#footnote-ref-10)