

استخراج ویژگی با شبکه‌های مولد تخاصمی عمیق^۱ و رمزگذار خودکار پیچشی^۲

حامد محمدزاده

(۱) چکیده

(۲) مقدمه

در این سری آزمایشات از مدل‌های AutoEncoder و GAN برای یادگیری بازنمایی بدون سوپریویژن استفاده می‌شود. در فاز اول مدل‌ها و دقت طبقه بند بر داده CIFAR10 تست می‌شود و در فاز دوم بر روی داده‌های چالش برانگیزتر DomainNet مدل‌ها آموزش داده می‌شوند. نتایج دقت‌ها و تصاویر خروجی مدل‌ها در هر بخش گزارش شده است.

(۳) انتخاب مدل

۱. AE مدل

برای مدل AE، در بخش Encoder این مدل پنج بلاک اصلی وجود دارد که در هر بلاک به جز بلاک آخر از دو لایه پیچشی، Batch Norm و Average Pooling استفاده شده. از بلاک اول تا پنجم به ترتیب تعداد چنل‌ها [32, 64, 128, 256, 512] است. دقت شود که ابعاد خروجی Decoder ورودی خود که بردار 512 بعدی است را در طی پنج بلاک به ابعاد تصاویر ورودی که $32 \times 32 \times 3$ است می‌رساند. در هر بلاک Decoder از یک لایه پیچشی ترانهاده^۳، Batch Norm و relu استفاده شده. (به غیر از لایه آخر که از Sigomid استفاده می‌کند) مدل استفاده شده برای داده‌های DomainNet با کمی تغییر تصاویر $64 \times 64 \times 3$ را به عنوان ورودی می‌گیرد و آن‌ها را بازسازی می‌کند.

۲. GAN مدل

برای انتخاب معماری این مدل از [۴] و پیاده‌سازی آن در [این لینک](#) الهام گرفته شده. بخش مولد از فضایی 1×100 طی پنج بلاک اصلی به تصویری با ابعاد $32 \times 32 \times 3$ می‌رسد، در هر بلاک تعداد چنل‌ها به ترتیب [64, 128, 256, 512] است که از یک لایه پیچشی ترانهاده، Batch Norm و relu استفاده شده. بخش Discriminator شامل پنج بلاک از لایه‌های پیچشی و Batch Norm است. علاوه بر این

Deep Convolutional Generative Adversarial Networks^۱
Convolutional AutoEncoders^۲
ConvTranspose^۳

طبق مدل استفاده شده در [1] ، لایه‌های پیچشی و پیجشی ترازهاده بدون bias هستند وزن‌های اولیه‌شان از توزیعی نورمال با میانگین 0.0 و انحراف معیار 0.02 محسابه می‌شود. برای داده‌های DomainNet هم مدلی مشابه با کمی تغییرات تا بتواند تصاویر 64×3 را تولید کند استفاده می‌شود.

۳. طبقه‌بند کلاس‌ها

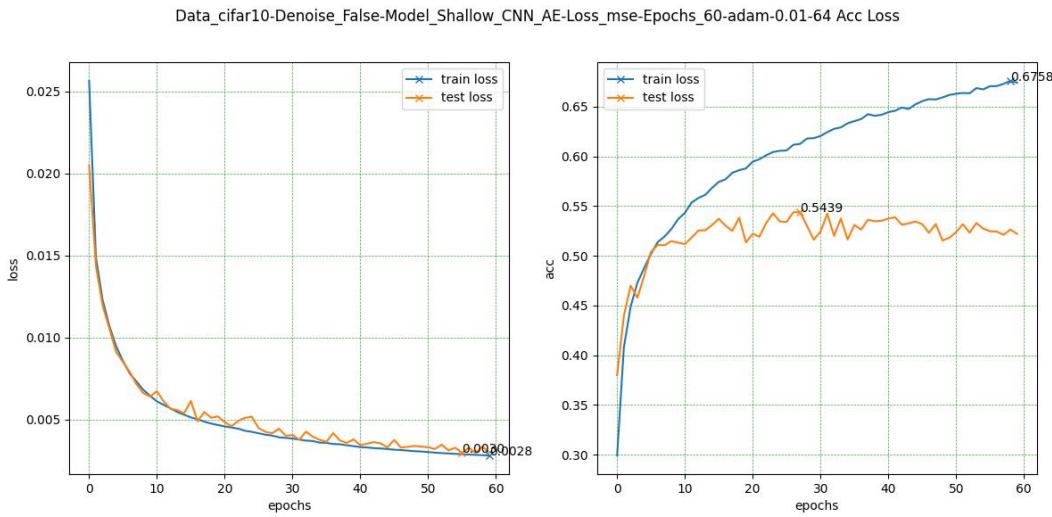
برای طبقه‌بندی کلاس‌ها با استفاده از بازنمایی‌های تولید شده از دو مدل ذکر شده، از مدلی با یک لایه مخفی به اندازه 128 استفاده شده است. برای استفاده این طبقه‌بند بر بازنمایی‌های مدل AE ، از خروجی Encoder که بردارهایی به ابعاد ۵۱۲ است استفاده می‌شود. برای مدل GAN هم بازنمایی قبل لایه آخر Discriminator با ابعاد ۴۰۹۶ استفاده می‌شوند.

۴) فاز اول: داده CIFAR10

در این بخش نتایج دو مدل بر داده‌های CIFAR10 را بررسی می‌کنیم.

۱. نتایج AE

مدل را ۶۰ مرحله بر داده‌های CIFAR10 با استفاده از batch size با مقدار ۱۲۸ آموزش می‌دهیم. تصاویر بازسازی شده از داده‌های آموزشی و داده‌های تست بعد از ۶۰ مرحله آموزش در شکل ۱۰ و ۱۱ در ضمیمه آورده شده‌اند. شکل ۱ نمودار یادگیری این مدل در طی آموزش است، در طی آموزش از تابع زیان MSE استفاده شده که در سمت چپ شکل ۱ آورده شده‌است. همچنین شکل سمت راست نمودار دقت طبقه‌بند که از خروجی Encoder استفاده کرده است آورده شده است.

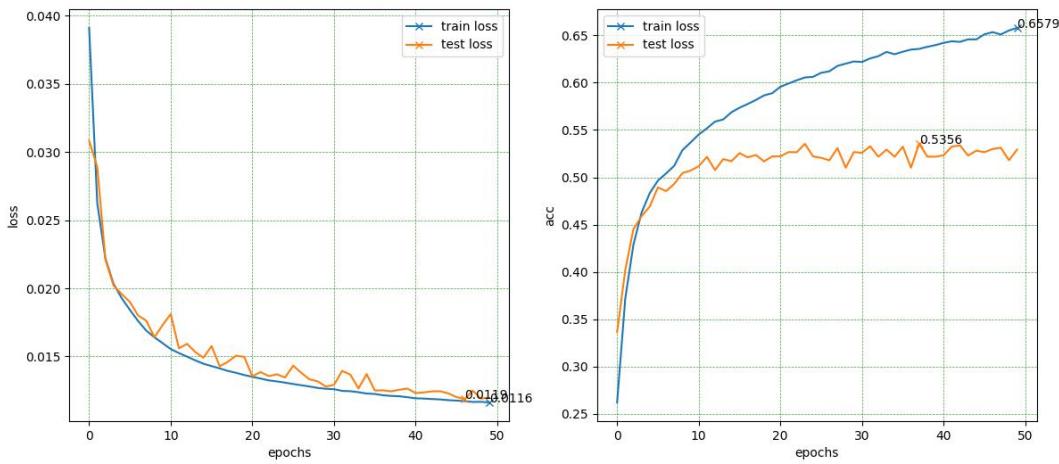


شکل ۱: نمودار یادگیری مدل AE بر داده‌های CIFAR10، شکل چپ نمودار تابع زیاد MSE برای تصاویر بازسازی شده و شکل راست نمودار دقت طبقه‌بندی که از بازنمایی‌های Encoder استفاده می‌کند است

۲.۴ مدل Denoising AE

مدل AE قسمت قبل با تصاویر نویزی هم آموزش دیده شده، به طوری که به تصاویر ورودی نویزی گاوین با ضریب ۰.۱ اعمال شد و تصاویر بدون نویز به عنوان هدف داده Decoder شد. خروجی مدل که نویز تصاویر ورودی را حذف کرده است در شکل ۱۲ و ۱۳ از ضمیمه آورده شده است. همچنان دقت طبقه‌بند با استفاده از بازنمایی‌های این مدل نسبت به حالت عادی بهبودی نیافت. شکل ۲ نمودار یادگیری مدل و دقت طبقه‌بند استفاده شده را نشان می‌دهد.

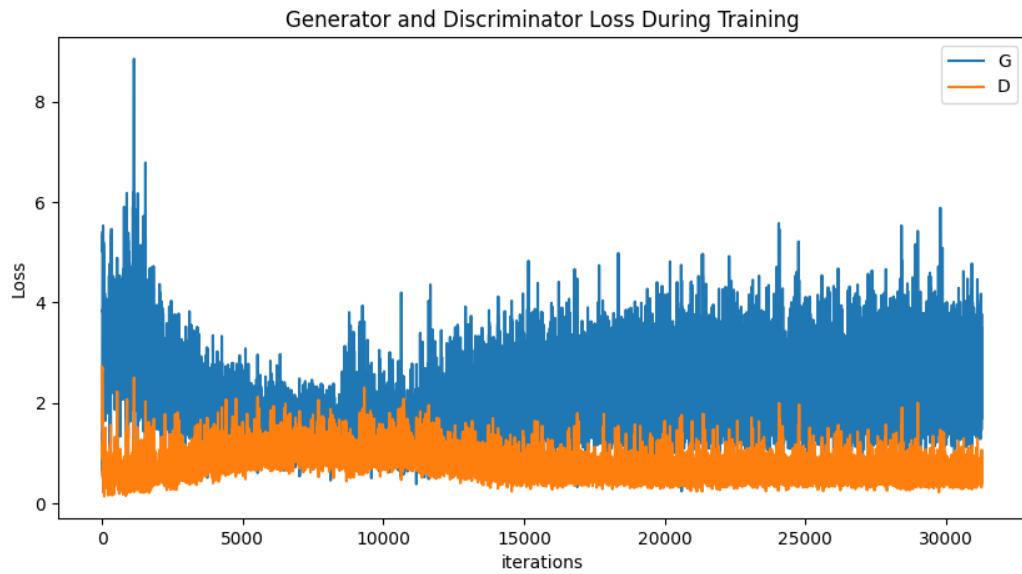
Data_cifar10-Denoise_True-Model_Shallow_CNN_AE-Loss_mse-EPOCHS_50-adam-0.01-128 Acc Loss



شکل ۲: نمودار یادگیری مدل AE با ورودی‌های نویزی بر داده‌های CIFAR۱۰، شکل چپ نمودار تابع زیاد MSE برای تصاویر بازسازی شده و شکل راست نمودار دقیق‌بندی که از بازنمایی‌های Encoder استفاده می‌کند است

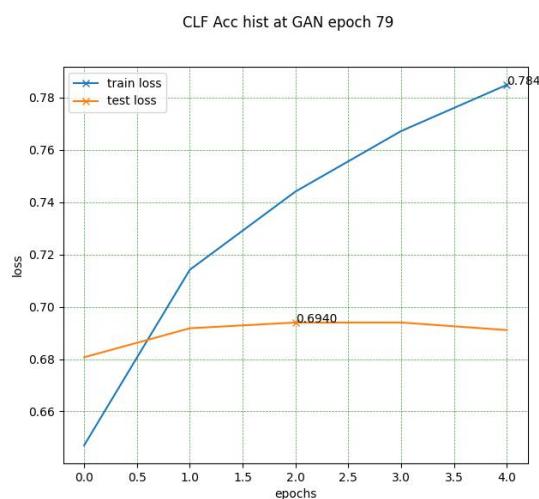
۳.۴ نتایج GAN

مدلی که در قسمت ۳.۱ برای داده‌های CIFAR۱۰ توضیح داده شد را ۸۰ مرحله با مقدار batch size ۱۲۸ آموزش می‌دهیم. مقدار توابع زیان شبکه مولد و شبکه Discriminator در هر گام از آموزش (هر batch) در شکل ۳ آورده شده است. طبق این شکل در مرحله ۲۲ از آموزش مقدار تابع زیان شبکه مولد به کمترین مقدار رسیده، تصاویر تولیدی غیرواقعی شبکه مولد در مرحله ۸۰ از یادگیری و مرحله ۲۲ از یادگیری در شکل ۱۴ و ۱۵ از ضمیمه آورده شده است.



شکل ۳: مقدار توابع زیان شبکه مولد و شبکه Discriminator در هر گام از آموزش

در این آزمایش طبقه بند تنها ۵ مرحله بر روی بازنمایی‌های شبکه Discriminator آموزش دیده است و نمودار یادگیری طبقه بند در این پنج مرحله در شکل ۴ آمده است.



شکل ۴: نمودار دقت طبقه بند در هر پنج مرحله از آموزش خود بر بازنمایی‌های مدل Discriminator شبکه که ۸۰ مرحله با داده Cifar10 آموزش دیده است

۴. نتایج فاز اول

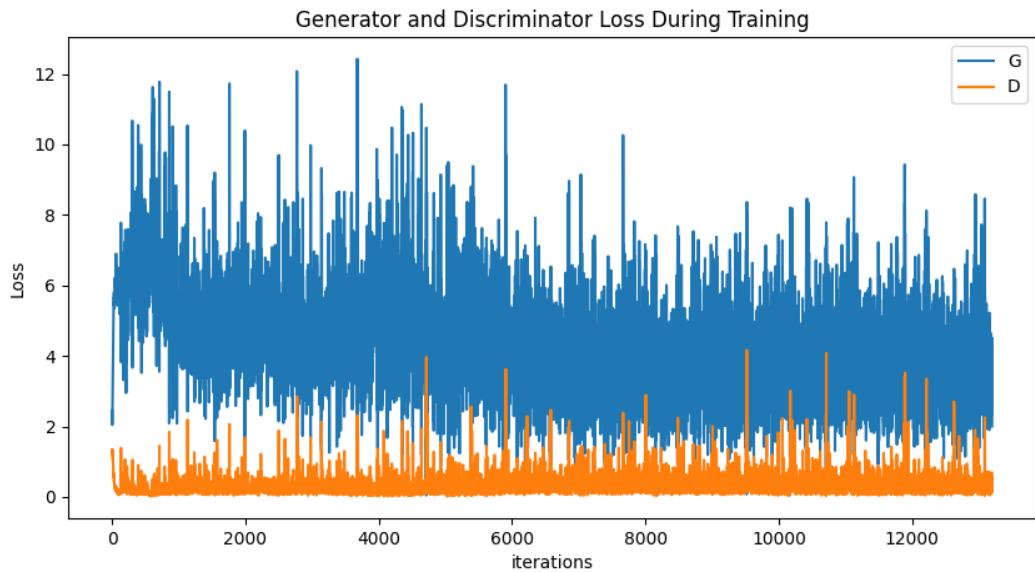
با مقایسه شکل ۲ و ۴ به نظر می‌آید که بازنمایی‌های بدستآمده از شبکه Discriminator از GAN برای طبقه‌بندی کلاس‌ها مناسب‌تر است. اما باید دقت شود که خروجی مدل Encoder ابعاد $1 \times 1 \times 512$ داشته باشد و موقعیت مکانی در آن حفظ نشده باشد، همنیطور ابعاد بازنمایی‌هایی که از GAN برای طبقه‌بندی استفاده می‌شود بسیار بیشتر از ابعاد خروجی Encoder است. (۴۰۹۶) همچنین مشاهده می‌شود که تصاویر بازسازی شده توسط AE تار تر از تصاویر اصلی است. دقت شود که ابعاد خروجی Encoder از ابعاد تصویر ورودی کمتر است و درواقع بازسازی Compression رخ داده است.

۵. فاز دوم: داده‌های DomainNet

از داده‌های DomainNet [۴] فقط ۱۶ کلاس مربوط به پستانداران را جدا می‌کنیم. تعداد داده‌ها و کلاس‌های انتخابی پستانداران دامین‌های real و painting در شکل ۱۶ در ضمیمه آمده است. همینطور همه تصاویر را به ابعاد 64×64 تغییر می‌دهیم. در تمامی این آزمایشات، منظور از داده آموزشی داده‌های مربوط به دامین real و منظور از داده‌های دامین painting می‌باشد. در مجموع حدود ۱۰ هزار داده آموزشی و ۵ هزار داده تست داریم.

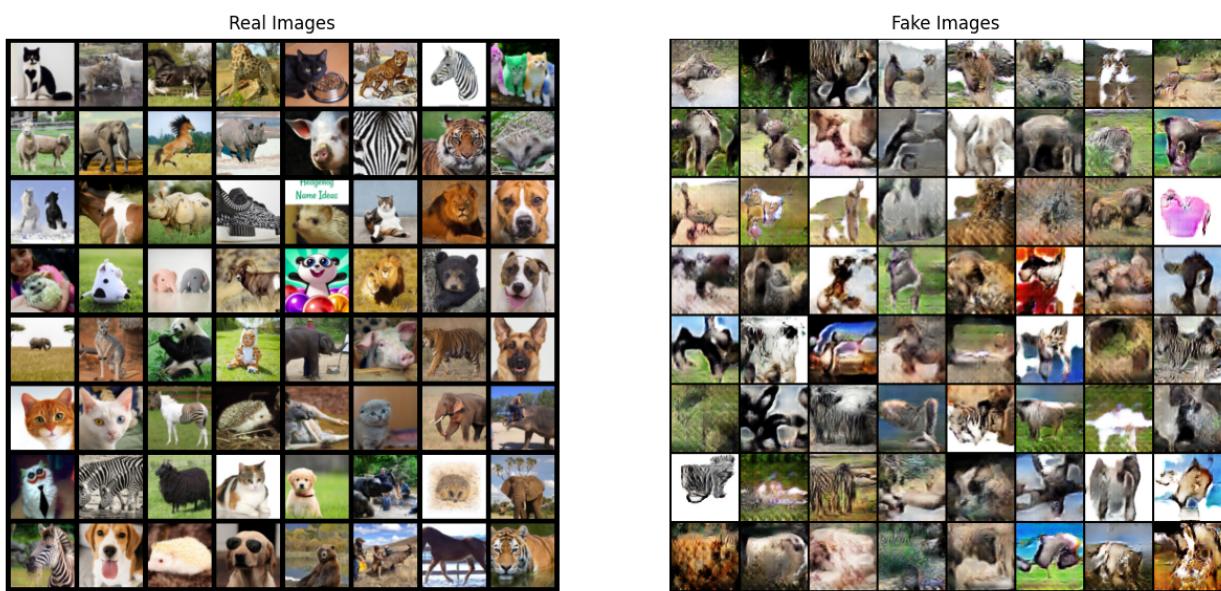
۱. نتایج GAN

در آزمایش‌های ابتدایی ای مدل توضیح داده شده در قسمت ۳.۲ مقدار تابع زیاد مولد با گذشت مراحل زیاد کم نمی‌شود و مقدار تابع زیان هم زیاد نمی‌شود، برای همین مدل Discriminator تغییری یافت بطوریکه کوچک تر شود و پیچیدگی کمتری داشته باشد به این امید که مولد بتواند در بازی MINMAX که با Discriminator دارد، برنده هم بشود. نمودار یادگیری مدل GAN بعد از این اصلاحات در شکل ۵ آمده است. لازم به ذکر است نمودار یادگیری مدل GAN قبل از اصلاحات ذکر شده در شکل ۱۷ در ضمیمه قرار دارد.



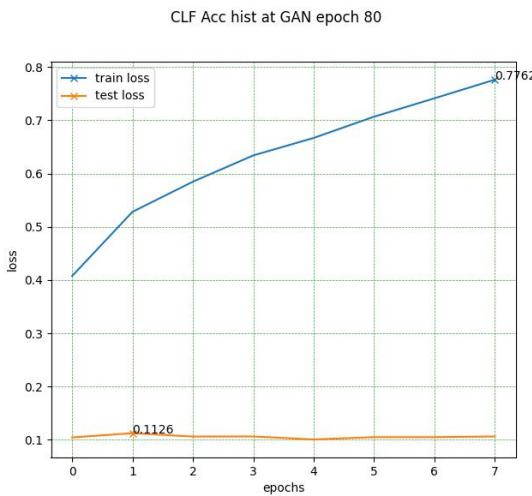
شکل ۵: مقدار توابع زیان شبکه مولد و شبکه Discriminator در هر گام از آموزش بر داده‌های DomainNet. معادل این نمودار برای قبیل از اصلاحات انجام شده در نمودار ۱۷ در ضمیمه قرار دارد. همچنین نموداری با جزئیات بیشتر درمورد یادگیری این مدل در شکل ۱۸ در ضمیمه آمده است

همچنین تصاویر غیرواقعی تولید شده توسط مولد در شکل ۶ آورده شده است. کیفیت تصاویر غیرواقعی تولید شده بالا است و بعضی از تصاویر ترکیبی عجیب از دو یا چند پستاندار است.



شکل ۶: شکل سمت راست تصاویر پستانداران واقعی داده‌های real از Domainnet است. شکل سمت چپ هم تصاویر غیر واقعی تولید شده توسط مولد است

در مرحله بعدی طبقه بند را ۸ مرحله برروی بازنمایی های داده های آموزشی (real) آموزش می دهیم. دقت این طبقه بند طی یادگیری در شکل ۷ آورده شده است.



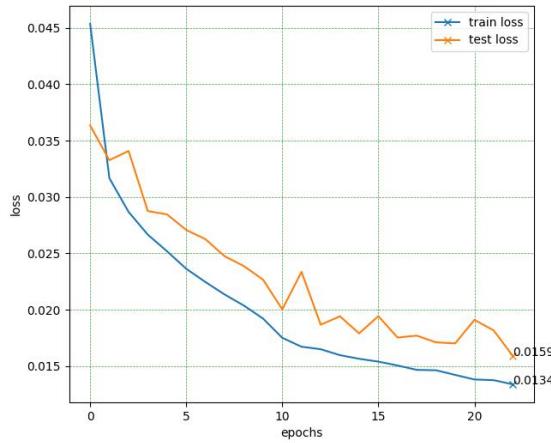
شکل ۷: نمودار دقت طبقه بند در هر ۸ مرحله از آموزش خود بر بازنمایی های مدل Discriminator شبکه که ۸۰ مرحله با داده real از domainnet دیده است. داده های تست مربوط به دامین painting هستند.

همانطور که در شکل ۷ واضح است، در بازنمایی هایی که طبقه بند یادمیگیرد تفاوت زیادی بین دامین real و painting وجود دارد و طبقه بند نمی تواند دقت خوبی از بازنمایی داده های painting بگیرد.

۲.۵ نتایج AE

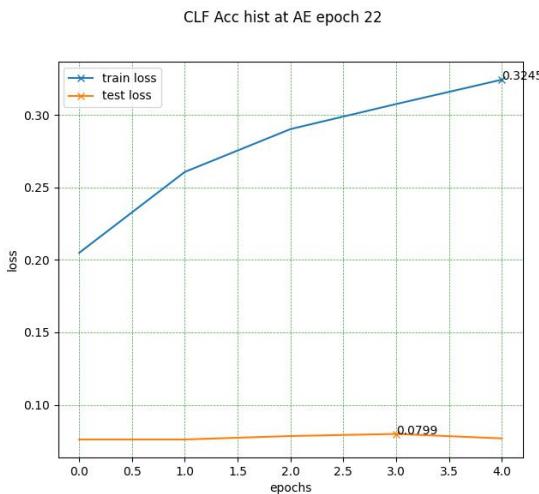
مدل توضیح داده شده در قسمت ۳.۲ را ۲۳ مرحله با داده های دامین real آموزش می دهیم. شکل ۸ نمودار تابع زیان MSE را برای مدل AE نشان می دهد. همچنین تصاویر بازسازی شده توسط AE در شکل ۱۹ درضمیمه آورده شده است.

mainnet-Denoise_False-Model_Shallow_CNN_AE-Loss_mse-Epochs_50-adam-0.01-64



شکل ۸: نمودار تابع زیان مدل AE برای داده‌های آموزشی و تست در طی یادگیری

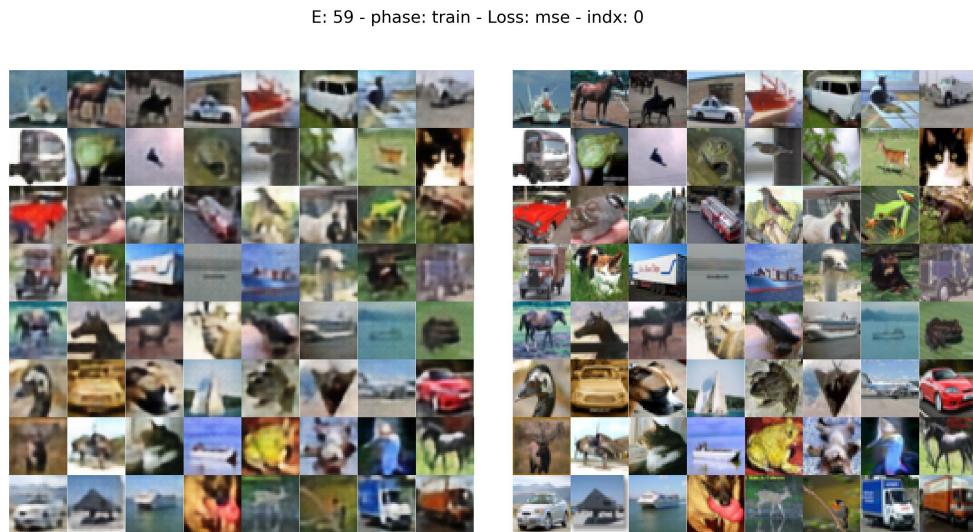
همچنین دقیق طبقه بند که با بازنمایی‌های داده‌های Real از دیکوادر آموزش داده شده در شکل ۹ آورده شده.



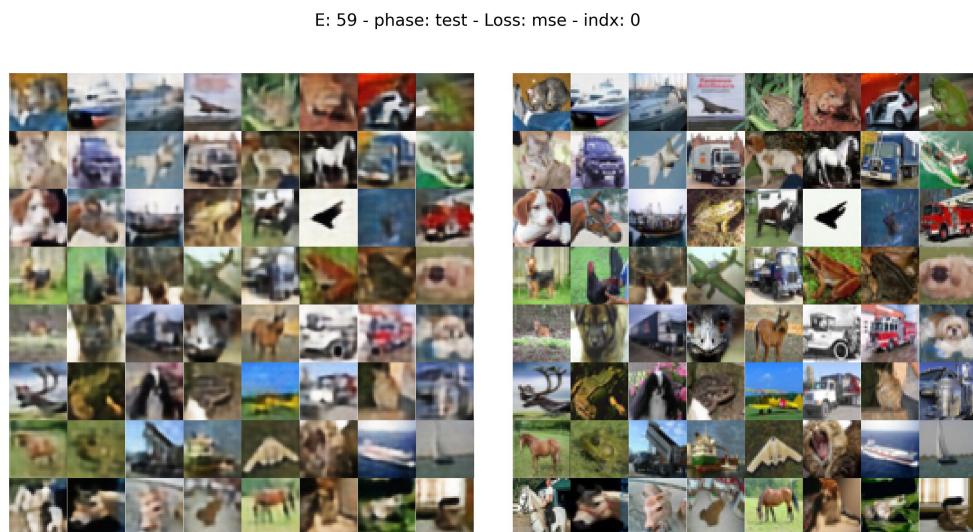
شکل ۹: دقیق طبقه بند که از بازنمایی انکوادر استفاده می‌کند برای داده‌های آموزشی (real) و داده‌های تست (painting).

مشاهده می‌شود که مثل مدل GAN، بازنمایی‌ها برای طبقه بندی یک دامین جدید مناسب نیست، هرچند که Decoder مدل AE می‌تواند تصاویر دامینی که ندیده را تا حد خوبی بازسازی کند. نمونه‌ای از این تصاویر در شکل ۲۰ ضمیمه آورده شده است.

(۶) ضمیمه

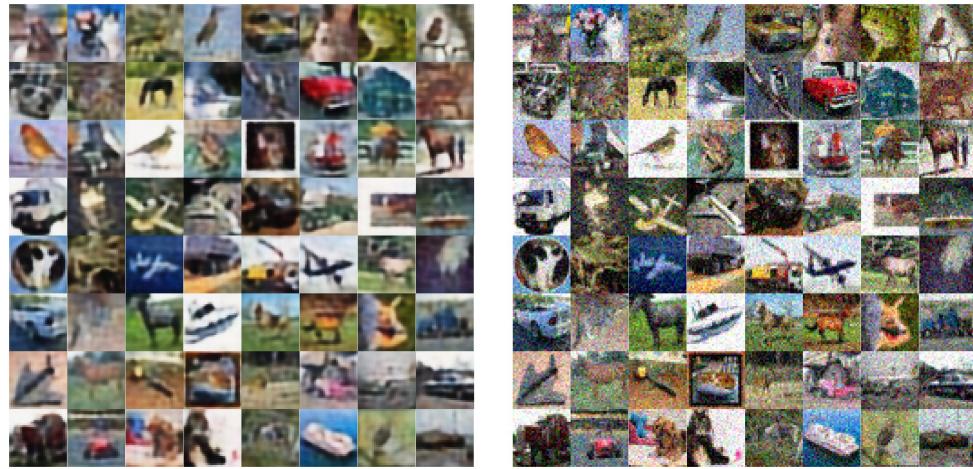


شکل ۱۰: خروجی داده‌های آموزشی برای مدل AE بعد از ۶۰ مرحله آموزش بر داده‌های CIFAR10. تصاویر ورودی در سمت راست و تصاویر بازسازی شده در شکل سمت چپ قرار دارند.



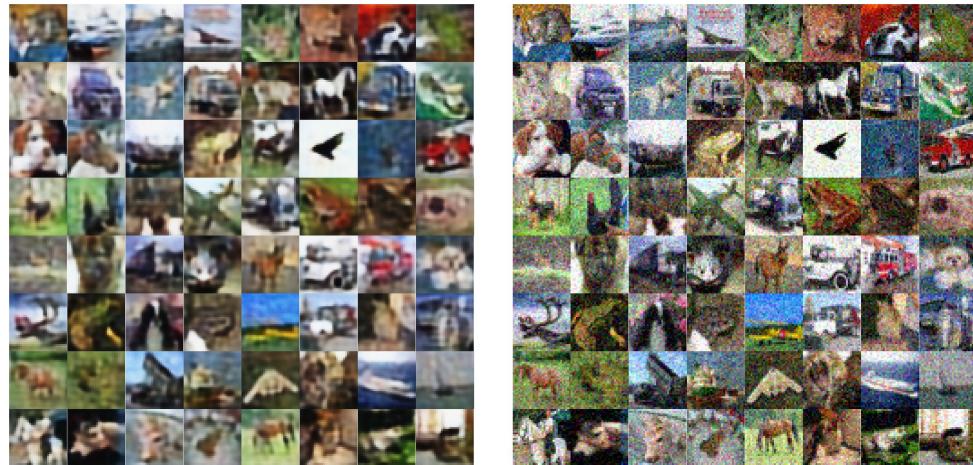
شکل ۱۱: خروجی داده‌های تست برای مدل AE بعد از ۶۰ مرحله آموزش بر داده‌های CIFAR10. تصاویر ورودی در سمت راست و تصاویر بازسازی شده در شکل سمت چپ قرار دارند.

E: 49 - phase: train - Loss: mse - indx: 0

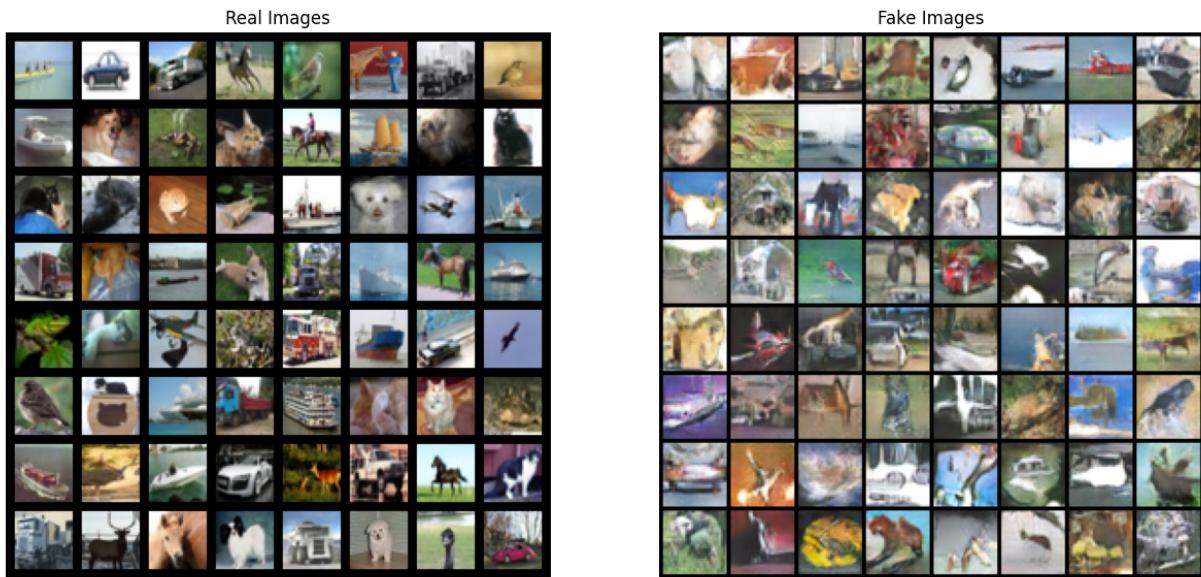


شکل ۱۲: خروجی داده‌های آموزشی برای مدل Denoising AE بعد از ۵۰ مرحله آموزش بر داده‌های CIFAR10. تصاویر ورودی در سمت راست و تصاویر بازسازی شده در شکل سمت چپ قرار دارند.

E: 49 - phase: test - Loss: mse - indx: 0



شکل ۱۳: خروجی داده‌های تست برای مدل Denoising AE بعد از ۵۰ مرحله آموزش بر داده‌های CIFAR10. تصاویر ورودی در سمت راست و تصاویر بازسازی شده در شکل سمت چپ قرار دارند.



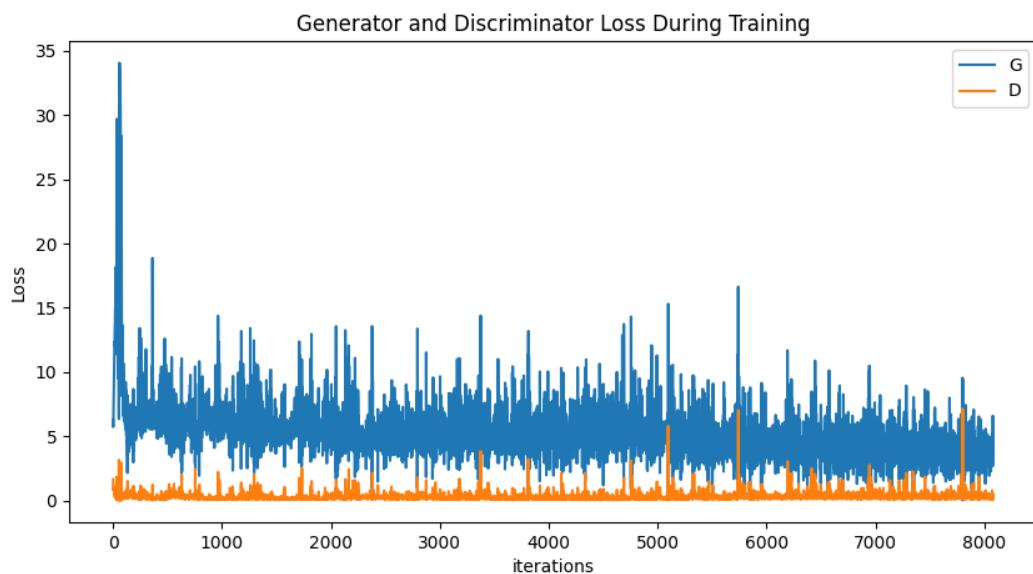
شکل ۱۴: در سمت چپ تصاویر واقعی داده CIFAR۱۰ و در سمت راست تصاویر تولیدی شبکه مولد قسمت ۴.۳ بعد از ۸۰ مرحله آموزش



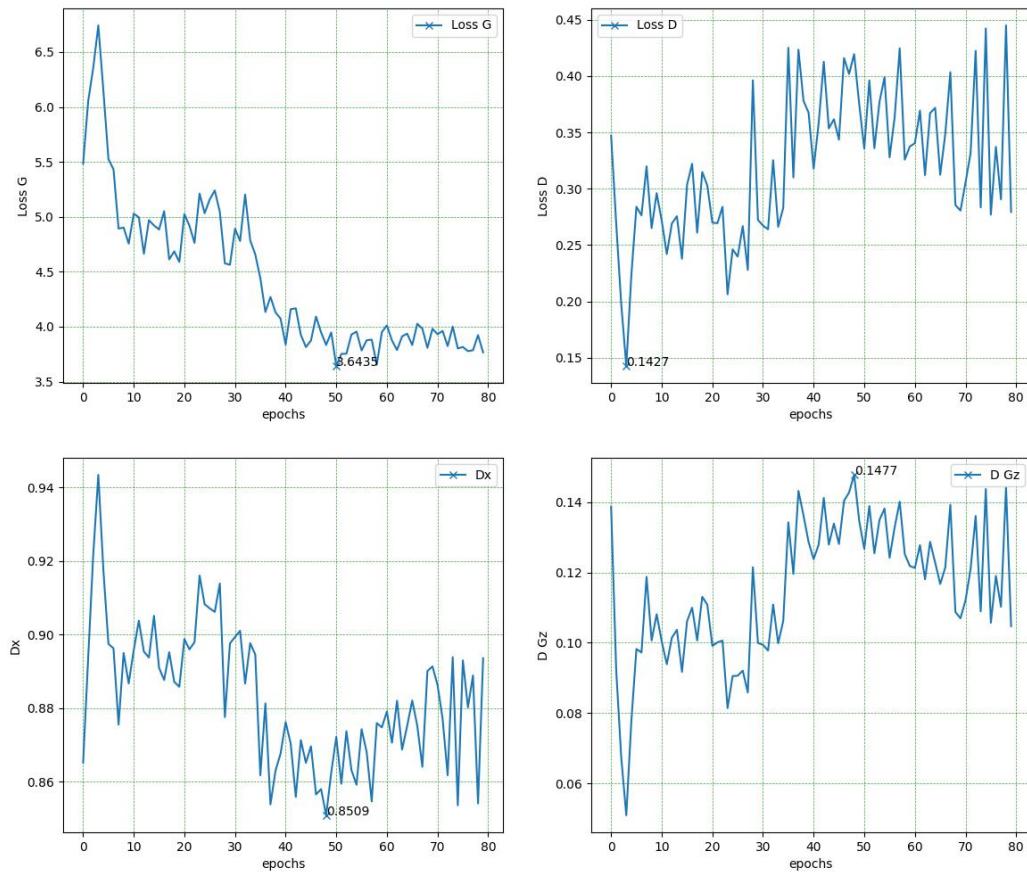
شکل ۱۵: در سمت چپ تصاویر واقعی داده CIFAR۱۰ و در سمت راست تصاویر تولیدی شبکه مولد قسمت ۴.۳ بعد از ۲۲ مرحله آموزش

Mammal													
bat	35	99	306	500	361	160	1461	bear	124	81	379	500	585
cat	43	172	344	500	796	130	1985	cow	188	134	156	500	541
dolphin	84	165	401	500	581	85	1816	elephant	115	188	425	500	789
hedgehog	138	48	248	500	727	109	1770	horse	201	216	521	500	645
lion	46	64	505	500	516	330	1961	monkey	123	85	405	500	699
panda	87	86	264	500	587	79	1603	pig	93	203	326	500	577
raccoon	187	24	249	500	676	348	1984	rhinoceros	102	91	220	500	684
squirrel	221	180	779	500	693	389	2762	tiger	315	285	422	500	607
zebra	235	306	298	500	683	278	2300						
								camel	154	31	289	500	493
								dog	70	225	721	500	782
								giraffe	134	172	105	500	594
								kangaroo	106	60	214	500	613
								mouse	74	50	445	500	147
								rabbit	105	135	269	500	695
								sheep	114	70	334	500	796
								whale	343	432	357	500	671

شکل ۱۶: تعداد داده‌ها و کلاس‌های پستانداران در دیتابیس DomainNet



شکل ۱۷: مقدار توابع زیان شبکه مولد و شبکه Discriminator قبل از اصلاحات ذکر شده در قسمت ۵.۱ در هر گام از آموزش



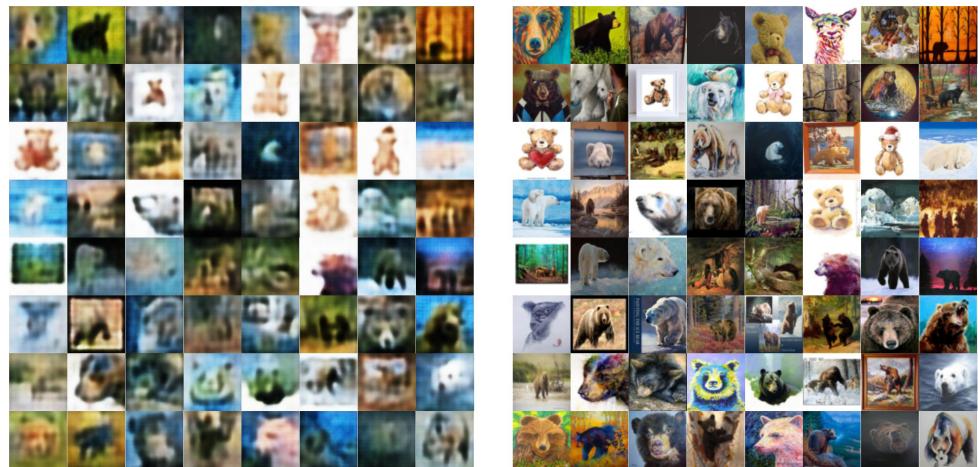
شکل ۱۸: جزییات آموزش مدل قسمت ۵.۱: دو نمودار بالایی تابع زیان مولد و Discriminator هستند (از چپ به راست). نمودار پایین سمت راست متوسط امتیاز تصاویر غیرواقعی تولید شده توسط مولد است که این امتیاز توسط Discriminator داده می‌شود، باید در طی آموزش به ۰.۵ نزدیک شود. نمودار پایین سمت چپ هم مقدار امتیاز داده‌های واقعی است که باید به ۰.۵ نزدیک شود

E: 43 - phase: train - Loss: mse - idx: 0



شکل ۱۹: خروجی داده‌های آموزشی برای مدل AE بعد از ۲۲ مرحله آموزش بر داده‌های real از DomainNet تصاویر ورودی در سمت راست و تصاویر بازسازی شده در شکل سمت چپ قرار دارند.

E: 42 - phase: test - Loss: mse - idx: 0



شکل ۲۰: خروجی داده‌های تست (داده‌های دامین painting که مدل آن را ندیده) برای مدل AE بعد از ۲۲ مرحله آموزش بر داده‌های real از DomainNet تصاویر ورودی در سمت راست و تصاویر بازسازی شده در شکل سمت چپ قرار دارند.

مراجع

convolutional deep with learning representation Unsupervised Chintala. Soumith and Metz, Luke Radford, Alec [i]

.٢٠١٦ networks, adversarial generative

multi-source for matching Moment Wang. Bo and Saenko, Kate Huang, Zijun Xia, Xide Bai, Qinxun Peng, Xingchao [j]

.٢٠١٩ adaptation, domain