

به نام خدا



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

درس شبکه‌های عصبی
استاد صفابخش

تمرین هفتم

علیرضا مازوچی
۴۰۰۱۳۱۰۷۵

سوال ۱

مدل GAN از دو بخش مولد (Generator) و متمایزگر (Discriminator) تشکیل شده است. بخش مولد هدفش ایجاد داده مشابه با داده‌های واقعی و بخش متمایزگر هدفش شناسایی و تفکیک داده‌های واقعی از داده‌های تقلبی است. مولد برای ساخت تصاویر تقلبی از داده‌های نویز استفاده می‌کند و در شبکه خود این نویز را تبدیل به یک تصویر با ابعاد و شرایط اولیه مشابه با تصاویر واقعی می‌کند. متمایزگر یک تصویر را می‌گیرد و میزان واقعی بودن آن را می‌سنجد. یعنی متمایزگر عملاً یک دسته‌بند است. تابع خطا در مدل GAN بدین شکل تعریف می‌شود که هر چه تصاویر تقلبی تولیدشده توسط مولد واقعی‌تر باشد و متمایزگر را به خطا بیاندازد به مولد هزینه کمتری داده می‌شود و هرچه متمایزگر بهتر بتواند تصاویر را تفکیک کند یعنی بفهمد تصاویر واقعی، واقعی هستند و تصاویر تقلبی، تقلبی هزینه کمتری برای آن در نظر گرفته می‌شود. این تابع به شرح زیر است:

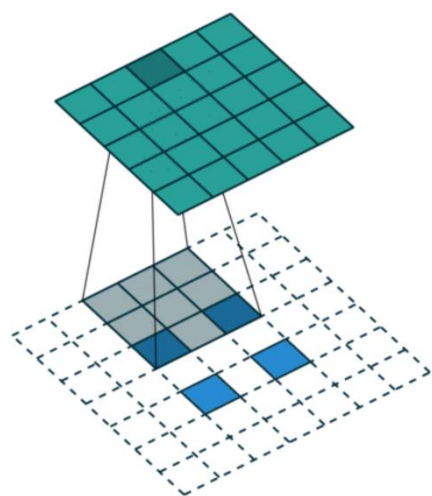
$$\min_{\theta_g} \max_{\theta_d} \left[\mathbb{E}_{x \sim p_{data}} \log D_{\theta_d}(x) + \mathbb{E}_{z \sim p(z)} \log(1 - D_{\theta_d}(G_{\theta_g}(z))) \right]$$

همانطور که در این تابع مشخص است، یک مدل سعی در بیشینه کردن تابع هزینه و یک مدل سعی در کمینه کردن آن دارند؛ به همین دلیل نوعی تقابل میان این دو مدل وجود دارد.

آموزش هر دو شبکه نیز همزمان انجام می‌گیرد. در ابتدا مولد تصاویر تصادفی ایجاد می‌کند؛ تصاویری که به راحتی از نوع واقعی قابل تفکیک است. به همین دلیل متمایزگر می‌تواند به آسانی آن را تشخیص دهد و مولد جریمه می‌شود. به واسطه این جریمه مولد سعی می‌کند تا تصاویر شبیه‌تر به واقعیت ایجاد کند. در همین حال که مولد قدرت‌مندتر می‌شود، متمایزگر نیز مجبور می‌شود تا ویژگی‌های مناسب‌تری از تصویر را استخراج کند. بین ترتیب متمایزگر هم قدرتمندتر می‌شود. هر چه مولد قوی‌تر شود، متمایزگر قوی‌تر می‌شود و هرچه متمایزگر قوی‌تر شود، مولد قوی‌تر. حاصل یک مولد با توان ایجاد داده‌های تقلبی با شباهت زیاد به واقعیت و یک متمایزگر با قدرت زیاد تشخیص تصاویر واقعی از تقلبی خواهد بود.

سوال ۲

لایه معکوس کانولوشن عملکردی مخالف لایه کانولوشن دارد؛ یعنی با دریافت یک نقشه ویژگی یک تصویر را می‌سازد. با نگاهی دیگر می‌توان دید که یک لایه کانولوشن اطلاعات پیکسل‌های مختلف را در یک پیکسل خلاصه می‌کند ولی یک لایه معکوس کانولوشن اطلاعات یک پیکسل را به پیکسل‌های مختلف گسترش می‌دهد.



برای ایجاد پیکسل جدید باید ورودی را حاشیه‌گذاری (Padding) کرد و سپس عمل کانولوشن را انجام داد. این بیت‌های حاشیه‌ای هم در میان پیکسل‌های ورودی و در روی مرزها قرار می‌گیرند. مانند تصویر روبرو:

سوال ۳

در سوال ۱ تابع هزینه کلی ارائه شده است. از روی این تابع می‌توان توابع هزینه را به صورت جداگانه برای هر یک از دو مدل ایجاد کرد. از آنجایی که هر مدل کنترل مستقیمی بر روی پارامترهای مدل دیگر ندارد، برای ایجاد توابع هزینه صرفاً باید توجه کرد که پارامتر مدل مذکور در کدام قسمت تابع هزینه کلی نقش داشته است. با این تفاسیر برای مدل متمایزگر تابع هزینه زیر بدست می‌آید:

$$\max_{\theta_d} \left[\mathbb{E}_{x \sim p_{data}} \log D_{\theta_d}(x) + \mathbb{E}_{z \sim p(z)} \log(1 - D_{\theta_d}(G_{\theta_g}(z))) \right]$$

برای مدل مولد نیز تابع هزینه زیر:

$$\min_{\theta_g} \mathbb{E}_{z \sim p(z)} \log(1 - D_{\theta_d}(G_{\theta_g}(z)))$$

تابع هزینه‌ای که برای مدل مولد پیشنهاد شده است می‌تواند بهینه‌تر شود و به فرم زیر در آید:

$$\min_{\theta_g} \mathbb{E}_{z \sim p(z)} \log(1 - D_{\theta_d}(G_{\theta_g}(z)))$$

این تابع از تابع قبلی از این جهت بهتر است که در اوایل آموزش مقدار گرادیان زیاد است و مدل مولد می‌تواند سریع‌تر آموزش ببیند.

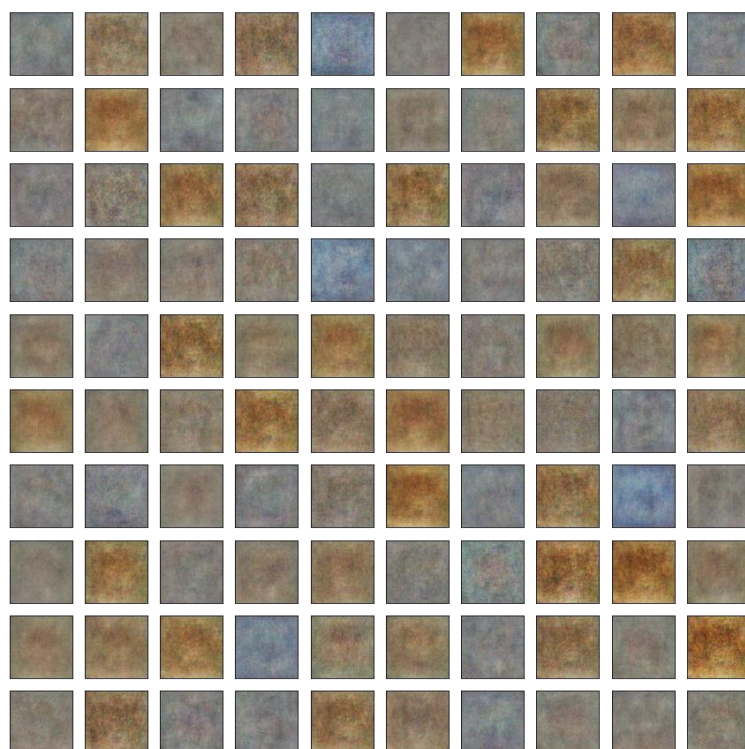
سوال ۴

ابتدا نکات کلی در مورد پیاده‌سازی را بیان می‌کنم:

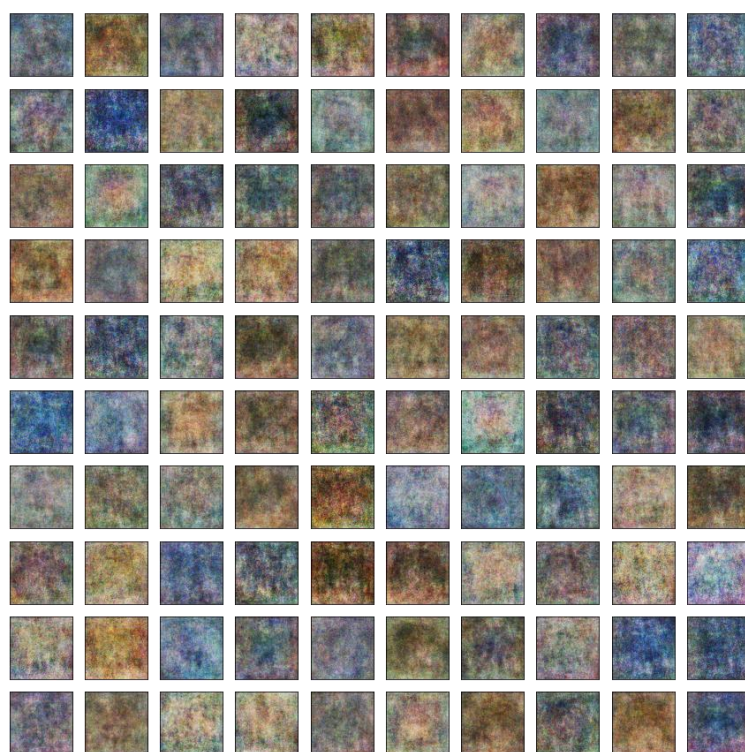
- برای پیاده‌سازی شبکه DCGAN و آموزش شبکه‌ها از معماری پیشنهادی موجود در اینترنت کمک گرفته‌ام^۱ و تنها آن را متناسب با مسئله بروز کرده‌ام.
- برای هر شبکه دو معماری توسعه داده‌ام: یک معماری معمولی با چندین لایه و یک معماری حداقلی با کمترین تعداد لایه.
- برای آموزش از ۲۵ گام استفاده کرده‌ام. قطعا آموزش بیشتر در مواردی می‌توانست نتایج بهتری را داشته باشد ولی با توجه به محدودیت‌های موجود به همین تعداد گام اکتفا کردیم.
- برای بررسی نتایج مدل مولد صد تصویر در گام ۵، ۱۵ و ۲۵ را نمایش می‌دهم. همچنین نمودار تغییرات مقدار خطای دو مدل در حین آموزش نشان داده شده است. جهت نمایش بهتر خطای تمایزگر پنج برابر نشان داده شده است.

در ادامه به ترتیب نتایج مربوط به FCGAN با معماری کوچک، FCGAN با معماری معمولی، DCGAN با معماری کوچک، DCGAN با معماری معمولی آورده شده است:

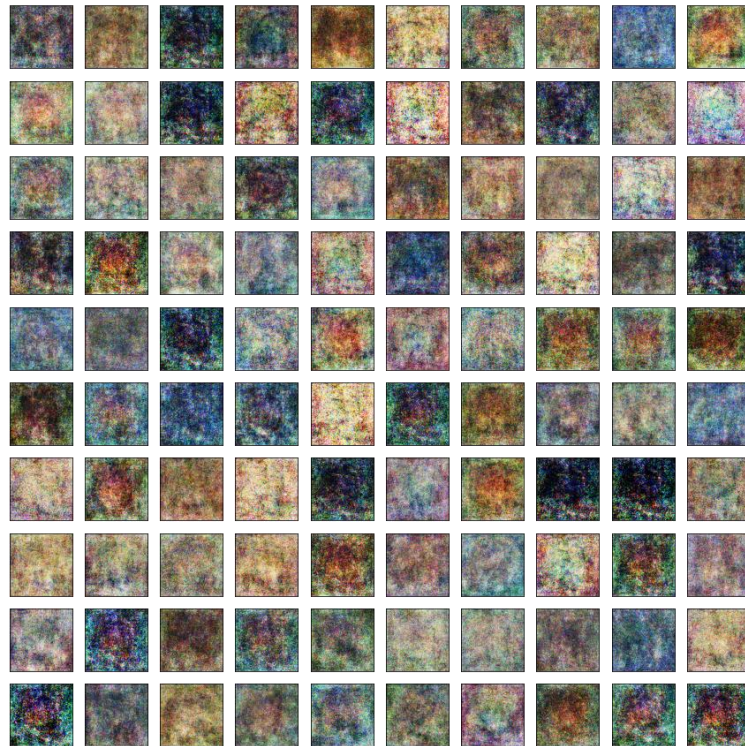
^۱ <https://www.tensorflow.org/tutorials/generative/dcgan>



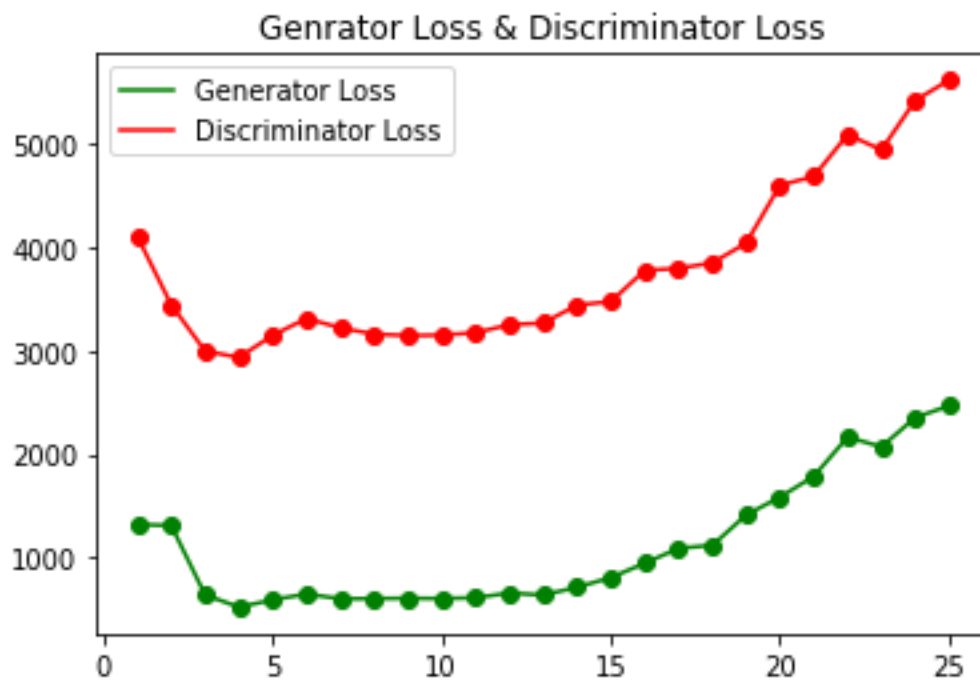
تصویر ۱ - تصاویر FCGAN کوچک در گام ۵



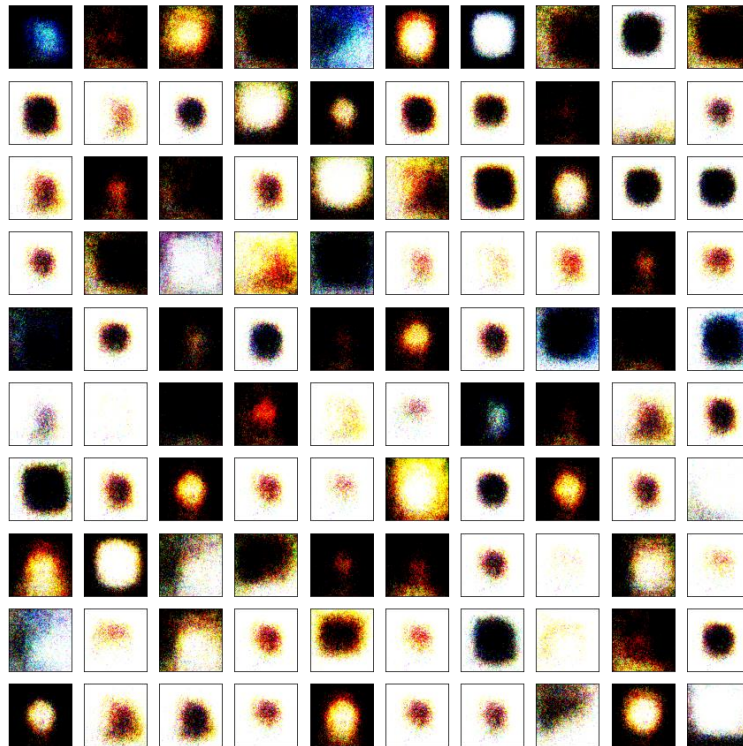
تصویر ۲ - تصاویر FCGAN کوچک در گام ۱۵



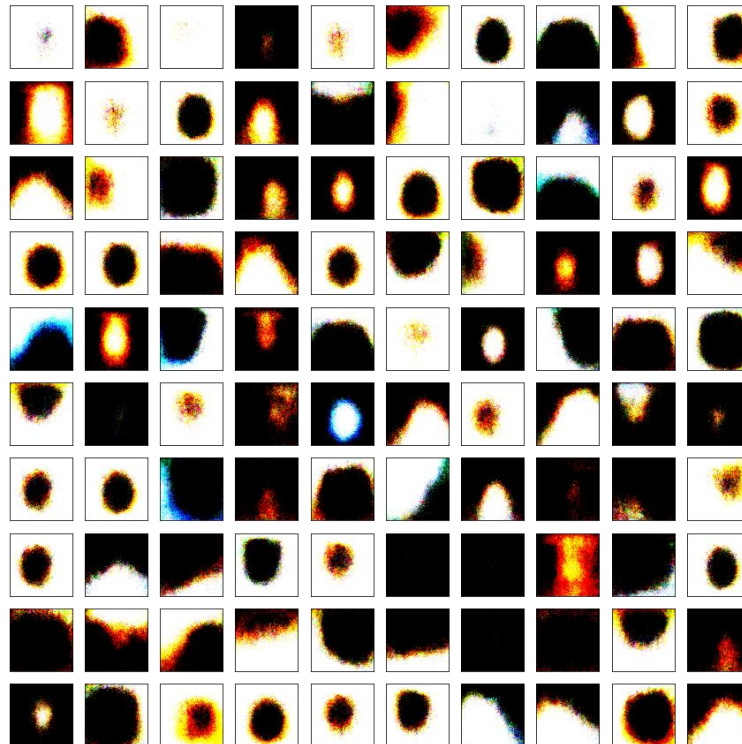
تصویر ۳ - تصاویر FCGAN کوچک در گام ۲۵



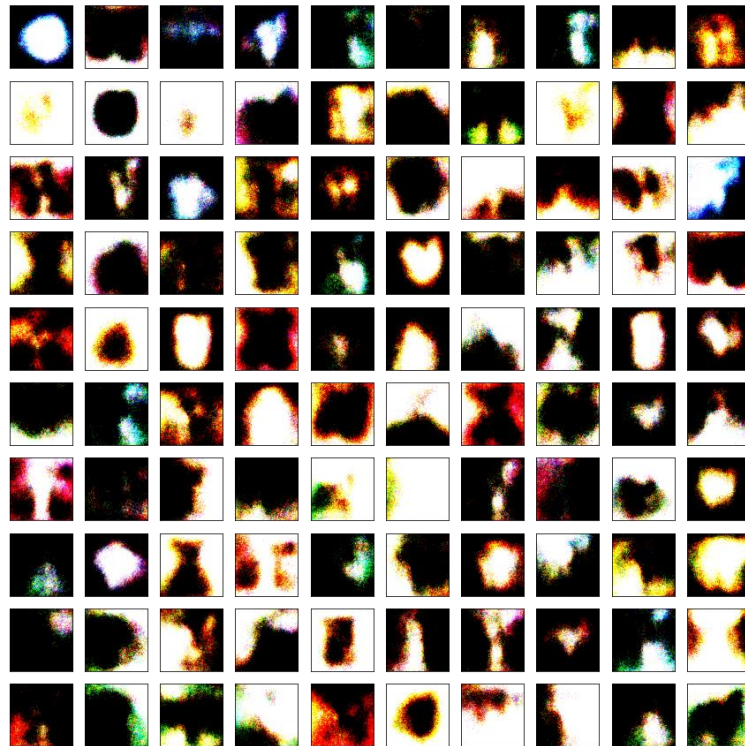
تصویر ۴ - تغییرات خطا در FCGAN کوچک



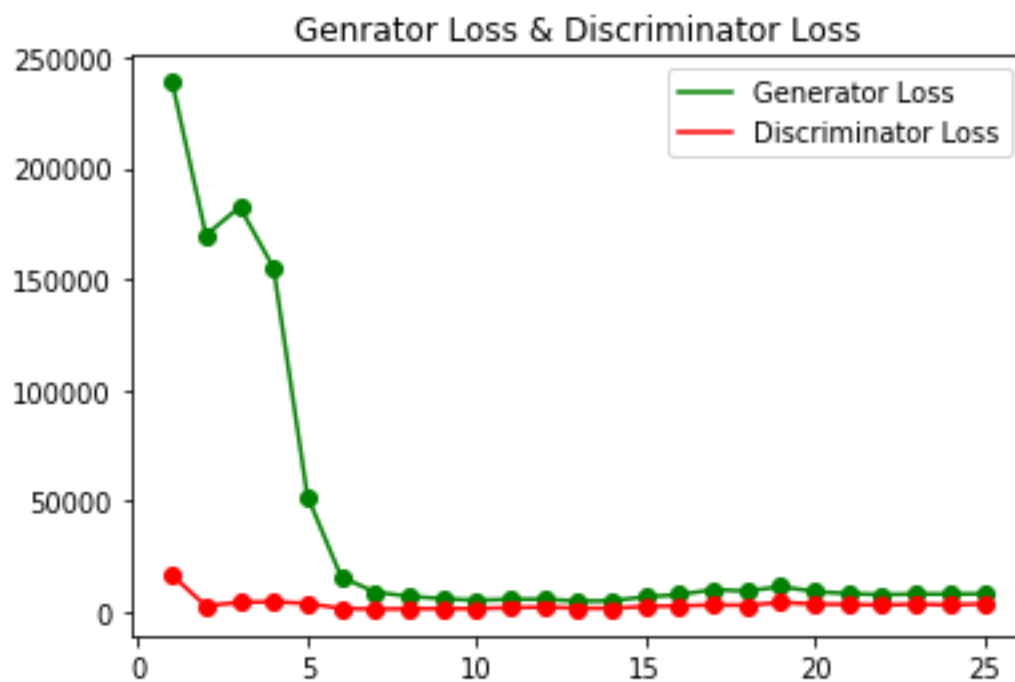
تصویر ۵ - تصاویر FCGAN معمولی در گام ۵



تصویر ۶ - تصاویر FCGAN معمولی در گام ۱۵



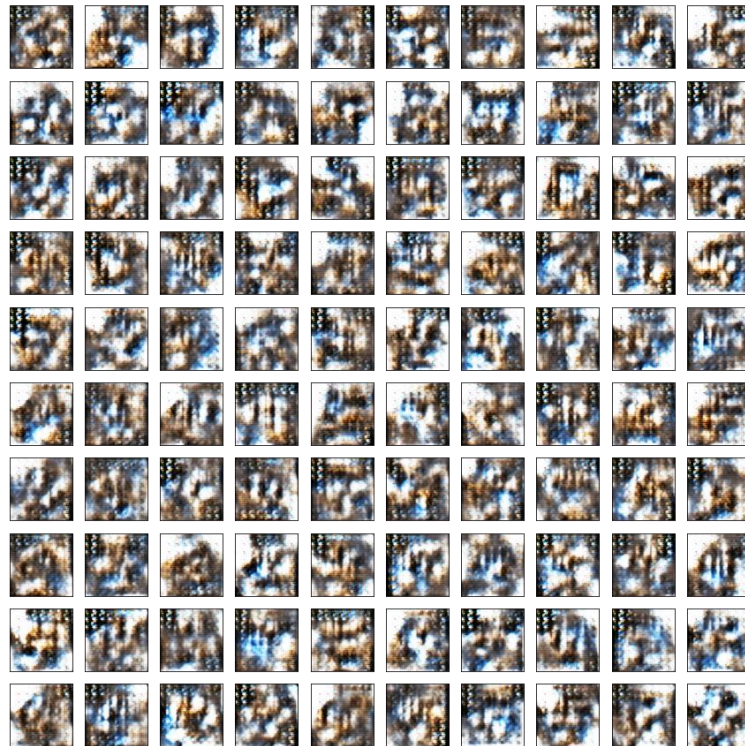
تصویر ۷ - تصاویر FCGAN معمولی در گام ۲۵



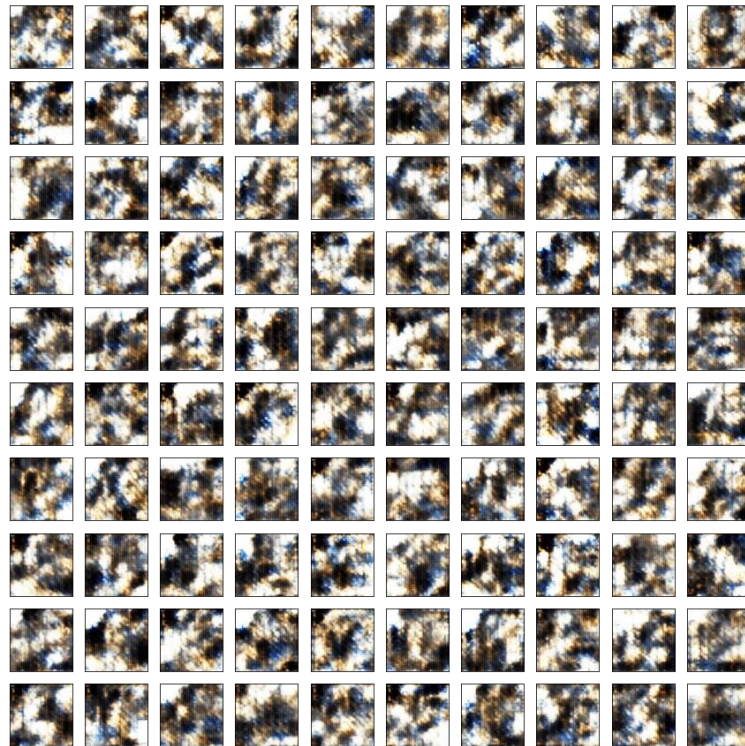
تصویر ۸ - تغییرات خطا در FCGAN معمولی



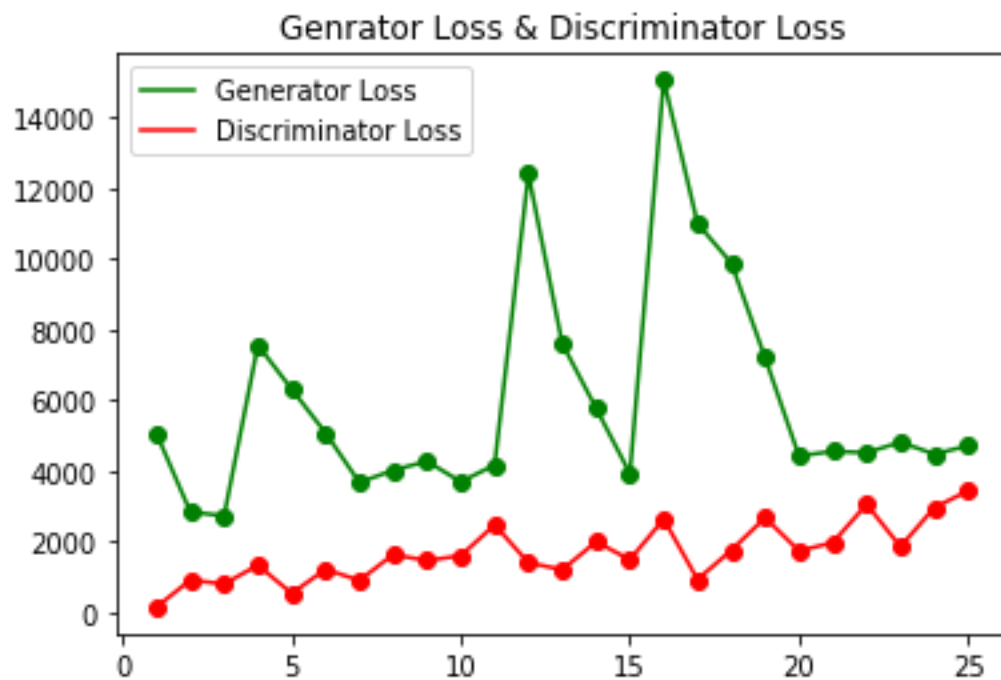
تصویر ۹ - تصاویر DCGAN کوچک در گام ۵



تصویر ۱۰ - تصاویر DCGAN کوچک در گام ۱۵



تصویر ۱۱ - تصاویر DCGAN کوچک در گام ۲۵



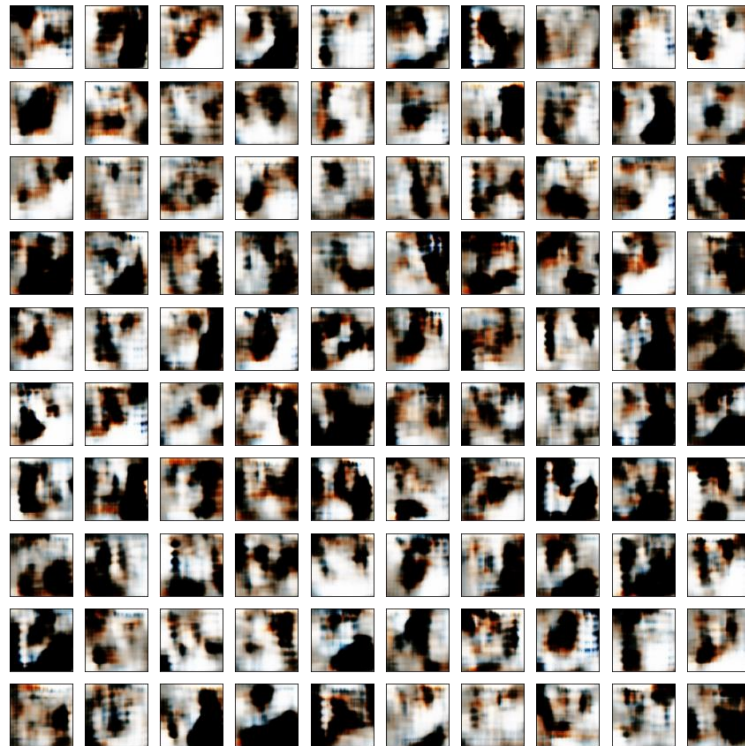
تصویر ۱۲ - تغییرات خطا در DCGAN کوچک



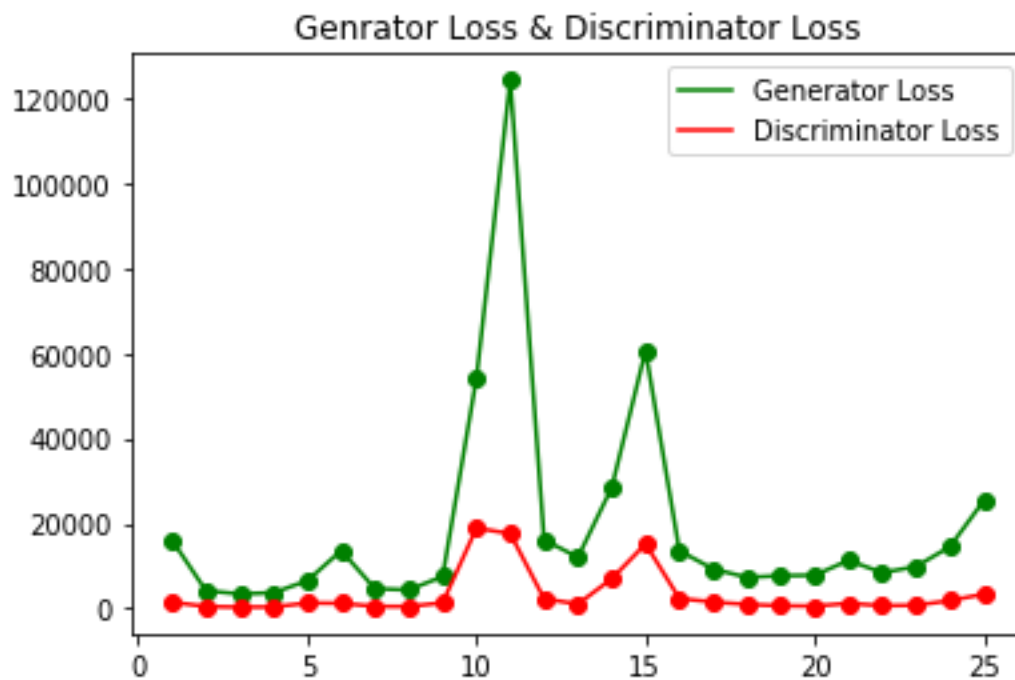
تصویر ۱۳ - تصاویر DCGAN معمولی در گام ۵



تصویر ۱۴ - تصاویر DCGAN معمولی در گام ۱۵



تصویر ۱۵ - تصاویر DCGAN معمولی در گام ۲۵



تصویر ۱۶ - تغییرات خطا در DCGAN معمولی

به طور کلی نتایج همه مدل‌ها با هر تعداد لایه مناسب نیست! ولی با این حال امکان مقایسه بین حالات مختلف هر شبکه وجود دارد.

در شبکه FCGAN با افزایش تعداد لایه‌ها خروجی بهتر شده است و تصاویر دارای جزئیات منطقی‌تری هستند. همچنین مشاهده می‌شود که خطای مولد شبکه با تعداد لایه کم‌تر به طور کلی در حال افزایش است. ولی در شبکه DCGAN مدل کوچک تصاویر نسبتاً بهتری را ایجاد کرده است.

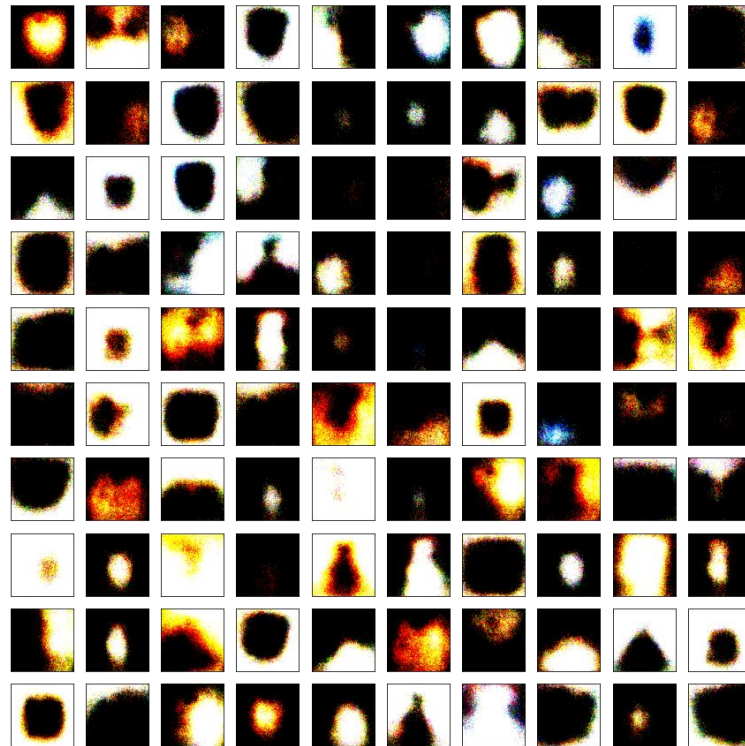
نهایتاً باید توجه داشت که در پیاده سازی من موقع استفاده از شبکه مولد کوچک از یک تمایزگر کوچک استفاده شده است. همین مسئله می‌تواند باعث شود تا خطای مولد برای شبکه‌های کوچک کم باشد ولی خروجی آن‌ها چندان جالب نباشد.

سوال ۵

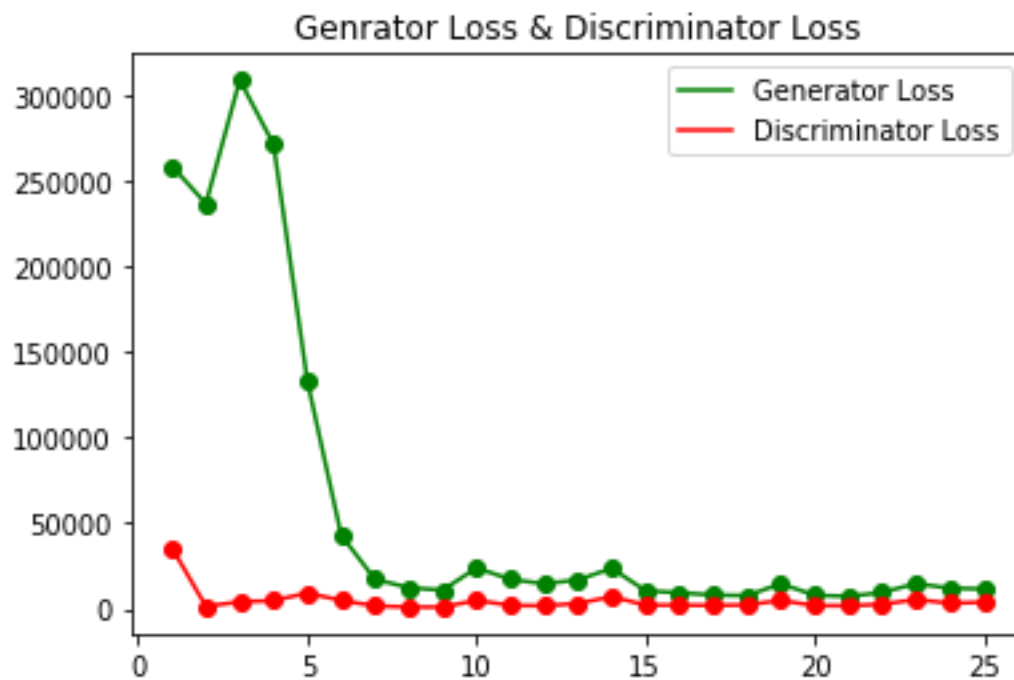
- یک راه آن است که تعداد تکرارهای مدل مولد را با مدل متمایزگر متفاوت بگیریم. یعنی به جای آن که یک بار مولد بروز شود و یک بار متمایزگر مثلاً یک بار مولد بروز شود و پنج بار متمایزگر.
- راه دیگر افزودن نویز به داده‌های واقعی است که در سوال ۶ ارائه شده است. مطابق با این راه و با افزودن نویز به تصاویر واقعی، مدل متمایزگر نمی‌تواند تصاویر واقعی را حفظ کند و مجبور می‌شود که ویژگی‌های اصلی آن را یاد بگیرد.
- می‌توان مقادیر نرخ یادگیری برای دو شبکه را متفاوت در نظر گرفت تا مدلی که همگرایی سریع‌تری دارد با سرعت کمتری به همگرایی نهایی برسد.

سوال ۶

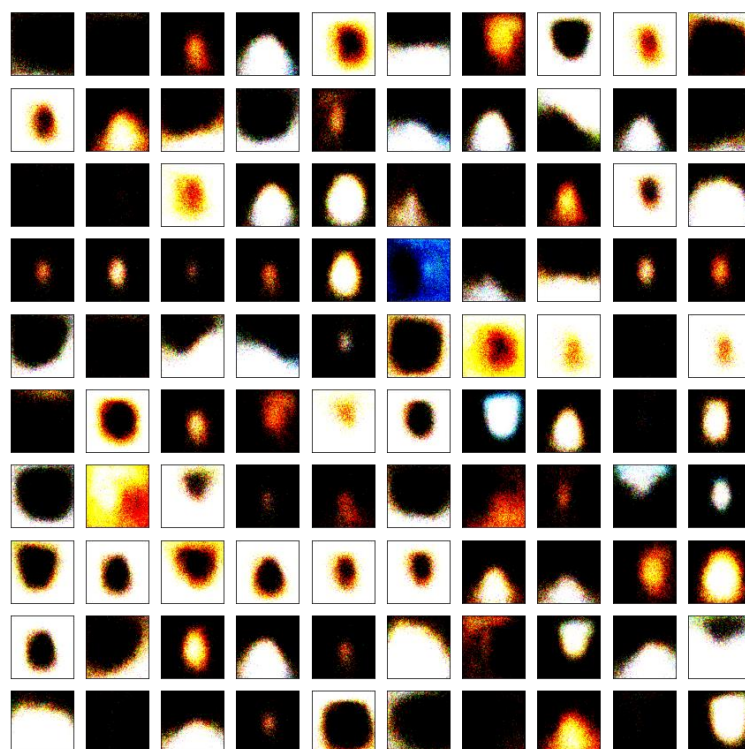
برای بررسی تاثیر نویز، نویز نرمال با میانگین صفر و انحراف معیارهای 0.1 ، 0.5 و 1 را بر روی تصاویر ورودی اعمال کردم و نتایج را جمع آوری کردم. برای این سوال و برای سادگی تنها تصاویر نهایی در گام آخر و نمودار تغییرات خطا را گزارش خواهم کرد:



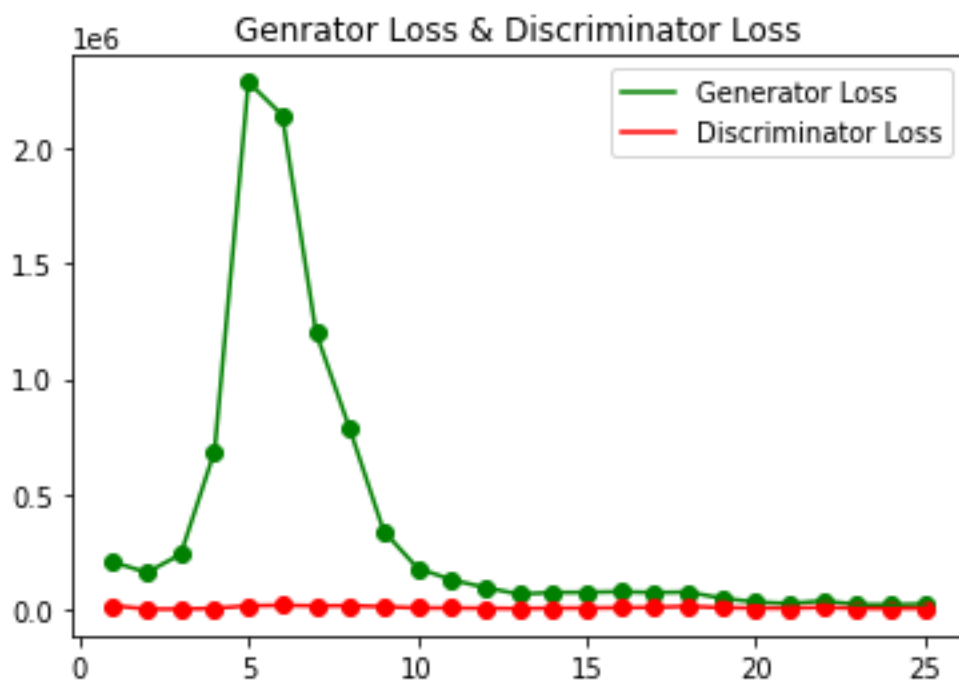
تصویر ۱۷ - تصاویر FCGAN با مقدار نویز ۰/۱



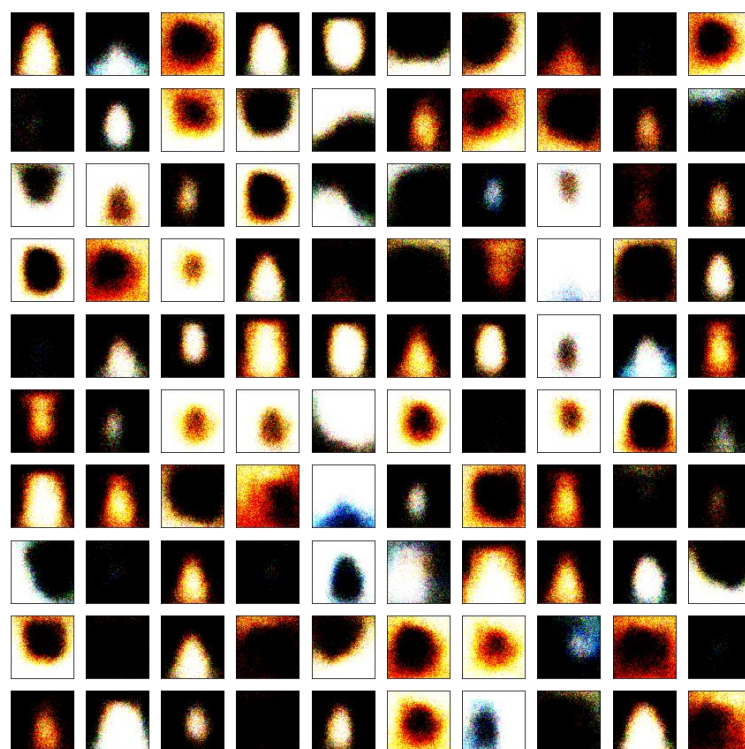
تصویر ۱۸ - تغییرات خطا FCGAN با مقدار نویز ۰/۱



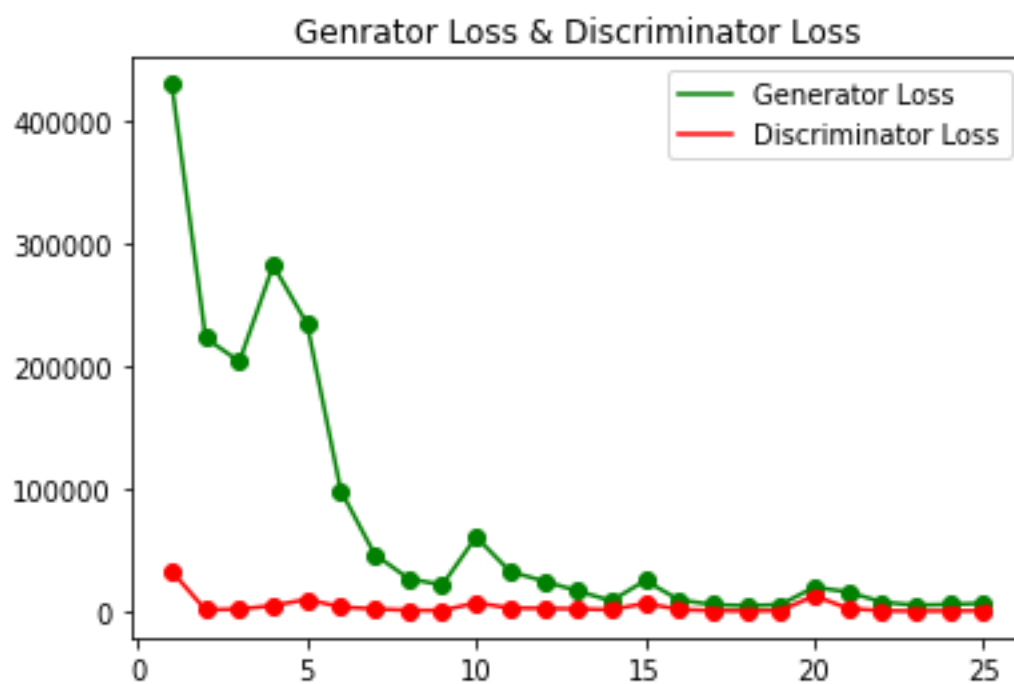
تصویر ۱۹ - تصاویر FCGAN با مقدار نویز ۰/۵



تصویر ۲۰ - تغییرات خطا FCGAN با مقدار نویز ۰/۵



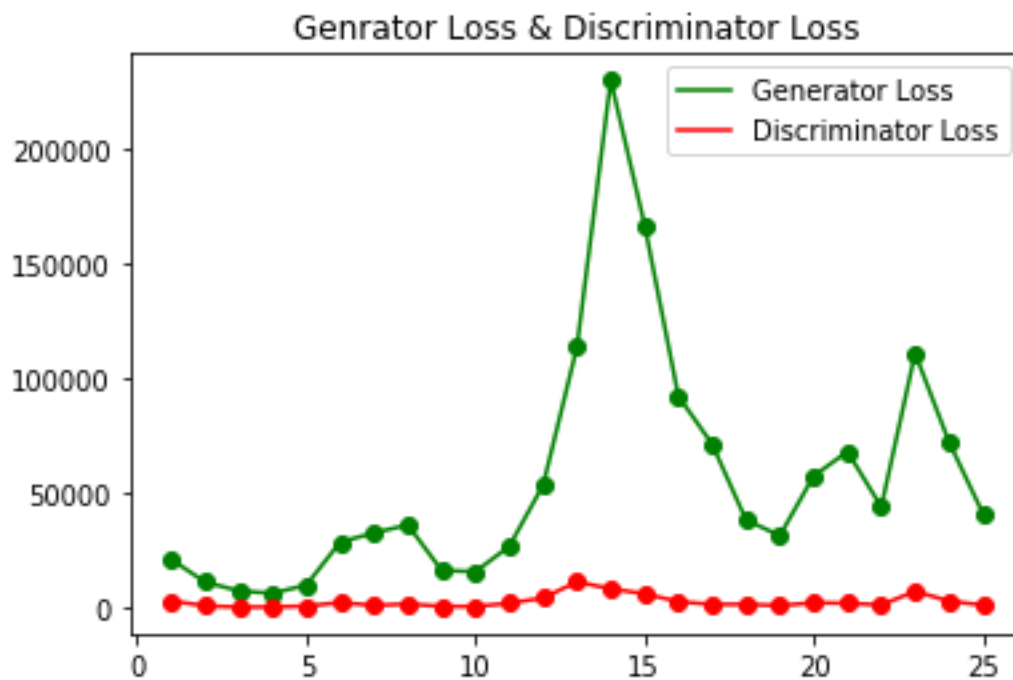
تصویر ۲۱ - تصاویر FCGAN با مقدار نویز ۱



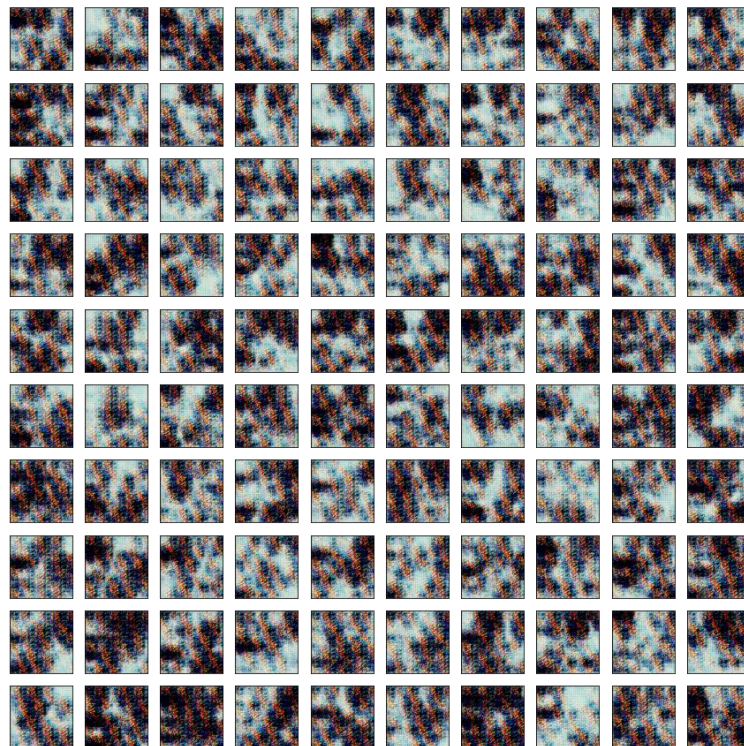
تصویر ۲۲ - تغییرات خطا FCGAN با مقدار نویز ۱



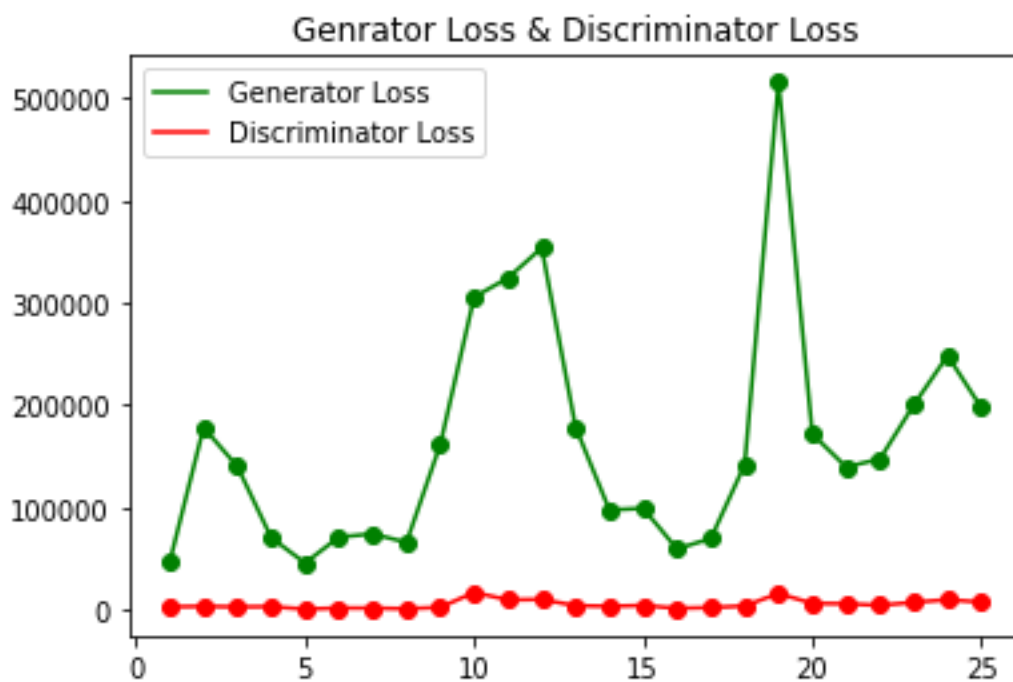
تصویر ۲۳ - تصاویر DCGAN با مقدار نویز ۰/۱



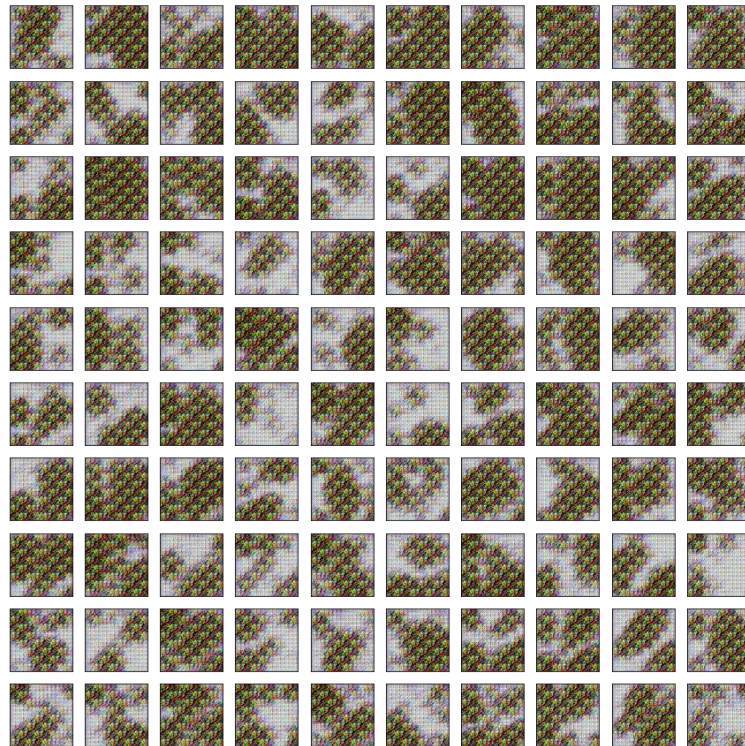
تصویر ۲۴ - تغییرات خطا DCGAN با مقدار نویز ۰/۱



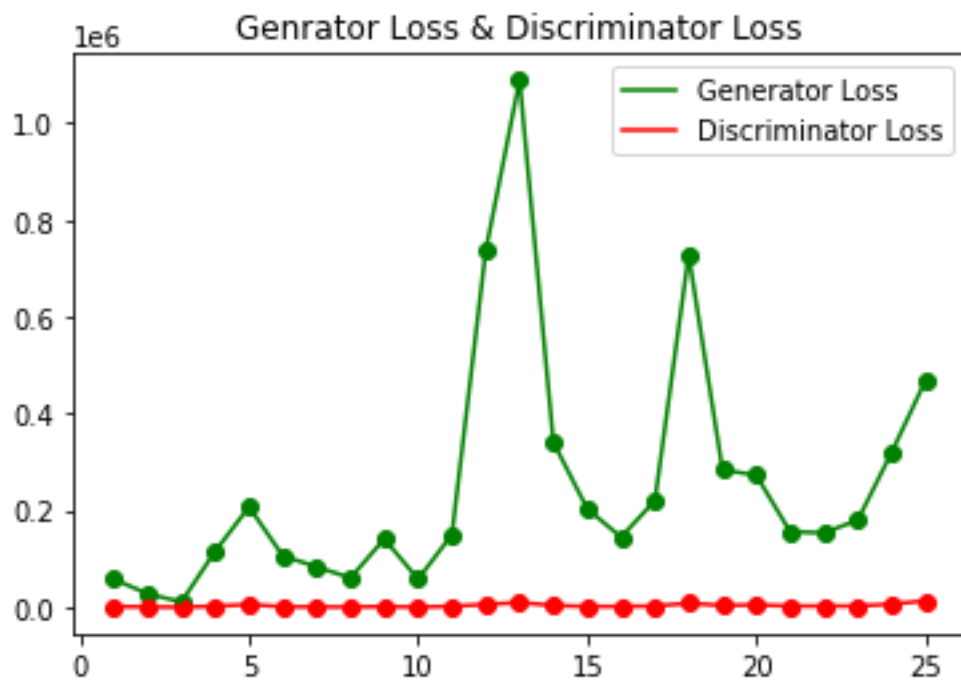
تصویر ۲۵ - تصاویر DCGAN با مقدار نویز ۰/۵



تصویر ۲۶ - تغییرات خطا DCGAN با مقدار نویز ۰/۵



تصویر ۲۷ - تصاویر DCGAN با مقدار نویز ۱



تصویر ۲۸ - تغییرات خطا DCGAN با مقدار نویز ۱

در مورد شبکه FCGAN وجود نویزها کمک خاصی به متعادل سازی نکرده است. کیفیت خروجی مولد هم تفاوت جدی ای ندارد. در شبکه DCGAN نیز نویز باعث افزایش شدید خطای مولد شده است و تصاویر اگر بدتر نشده باشد بهتر نشده است! بدین ترتیب نه تنها تعادل برقرار نشده است که شرایط بدتر هم شده است.

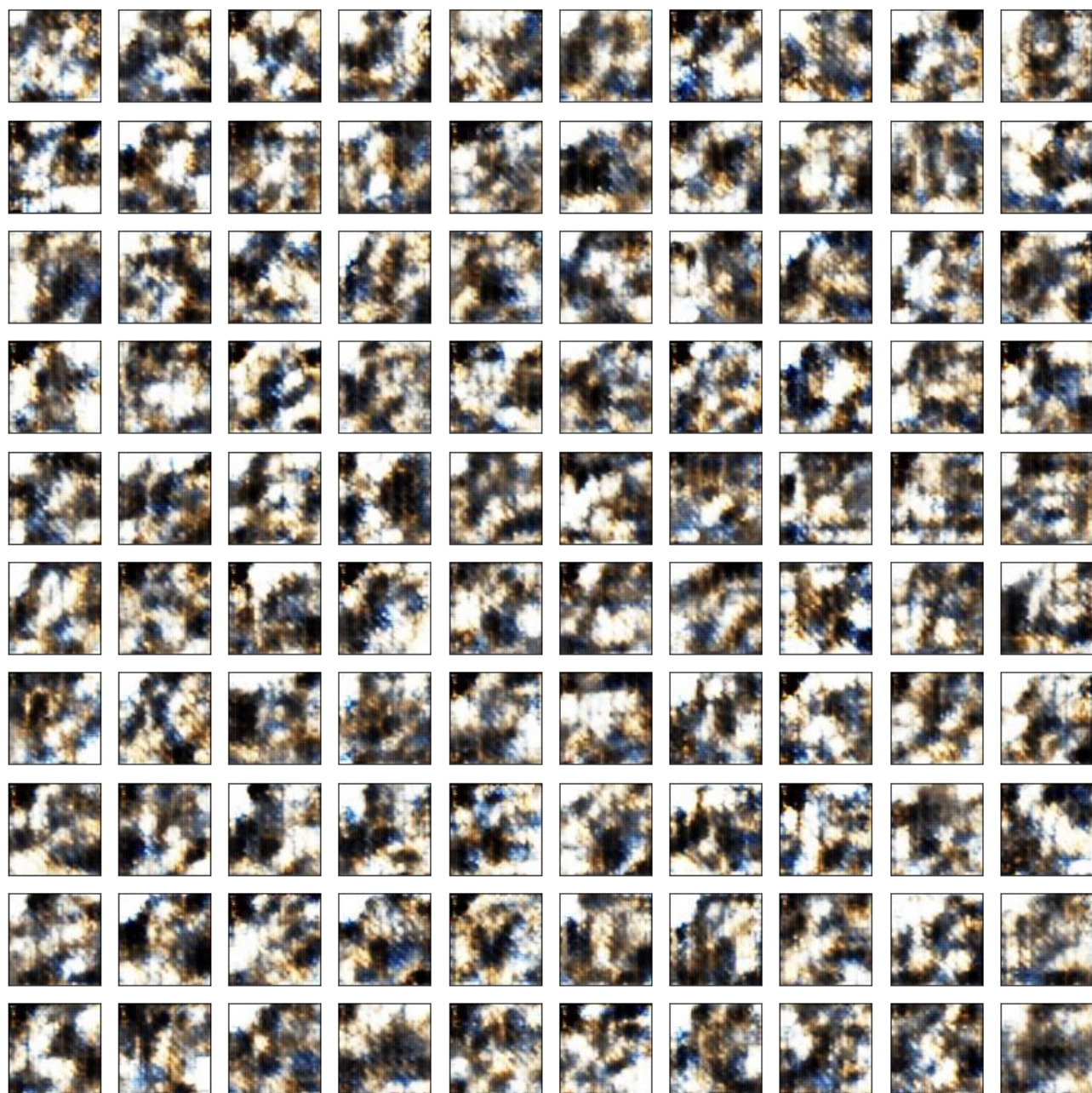
سوال ۷

شبکه DCGAN نتایج بهتری نسبت به شکل FCGAN داشته است. چراکه خروجی FCGAN غالباً از یک شبه دایره تشکیل شده است که الگوی واقعا ساده ای است اما خروجی DCGAN پیچیدگی بیشتری دارد و شباهت بیشتری به سگ و گربه دارد؛ هرچند که همچنان با تصاویر هدف بسیار فاصله دارد.

علت اینکه تصاویر DCGAN بهتر است به ساختار بهینه تر آنها بر می گردد. لایه های متراکم وزن های زیادی دارند و این باعث می شود در تعداد و ابعاد آنها محدودیت وجود داشته باشد ولی لایه های کانولوشنی به دلیل اشتراک وزن این مشکل را ندارند و امکان استفاده از تعداد لایه بیشتر آنها فراهم است. پس در شبکه های کانولوشنی با تعداد پارامتر برابر قدرت بیشتری برای تولید و ارزیابی تصویر وجود دارد.

سوال ۸

خروجی DCGAN کوچک که در قسمت های قبل ارائه شده است را مجدداً و با ابعاد بزرگ تر قرار می دهیم:



تصویر ۲۹ - خروجی مولد بهترین شبکه