



به نام خدا



دانشگاه تهران

دانشکده مهندسی برق و کامپیوتر

مدل‌های مولد عمیق

تمرین اول

سید محمد جزایری	نام و نام خانوادگی
810101399	شماره دانشجویی
1404/08/12	تاریخ ارسال گزارش

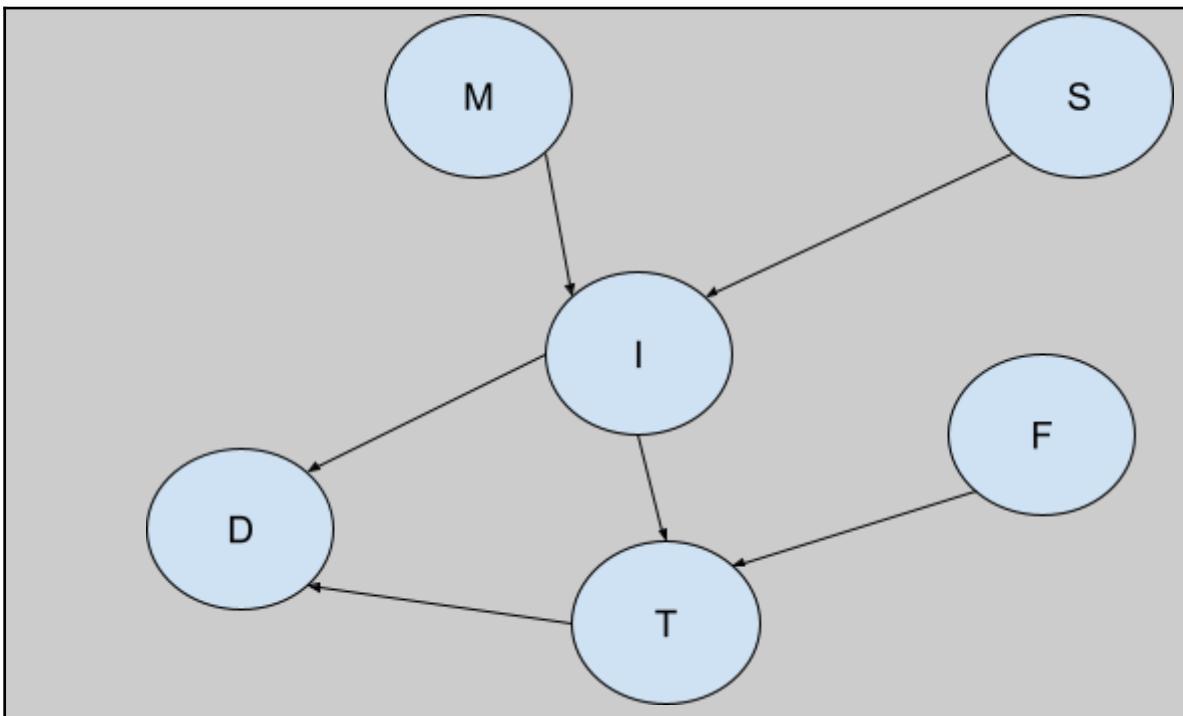
فهرست

3.....	سوال اول
3.....	بخش اول
3.....	زیر بخش اول
4.....	زیر بخش دوم
4.....	زیر بخش سوم
4.....	بخش دوم
4.....	زیر بخش اول
4.....	زیر بخش دوم
4.....	زیر بخش سوم
4.....	زیر بخش چهارم
5.....	زیر بخش پنجم
5.....	زیر بخش ششم
5.....	بخش سوم
5.....	زیر بخش اول
5.....	زیر بخش دوم
5.....	زیر بخش سوم
5.....	زیر بخش چهارم
5.....	بخش چهارم
6.....	سوال دوم
6.....	بخش اول
6.....	زیر بخش اول
6.....	زیر بخش دوم
6.....	زیر بخش سوم
6.....	زیر بخش چهارم
8.....	زیر بخش پنجم
8.....	زیر بخش ششم
10.....	زیر بخش هفتم
10.....	زیر بخش هشتم
11.....	بخش دوم
11.....	زیر بخش اول
11.....	زیر بخش دوم
12.....	زیر بخش سوم
13.....	مراجع

سؤال اول

بخش اول

زیر بخش اول)



(زیر بخش دوم)

$$P(M, S, I, F, T, D) = P(M)P(S)P(F)P(I|M, S)P(T|F)P(D|I, T)$$

(زیر بخش سوم)

- a. نادرست. F والد T است و T والد D پس یک مسیر بین آن ها وجود دارد.
- b. درست. هر مسیری که S را به D وصل می کند حتما از I می گذرد.
- c. درست. چرا که در مسیر T در راس M-I-T-F یک V-structure وجود دارد.
- d. نادرست. استدلال مانند بخش قبل است با معلوم شدن T بهم وابسته می شوند.
- e. نادرست. در مسیر D در راس M-I-D-T یک V-structure وجود دارد.

بخش دوم

(زیر بخش اول)

$$P(C, O, A, S, T, B, M) = P(C)P(O)P(A)P(S|O)P(T|O, A)P(B|S)P(M|B, T, A)$$

(زیر بخش دوم)

$$MB(T) = \{O, A, M, B\}$$

(زیر بخش سوم)

خیر چرا که A $\perp O$ در آن برقرار نیست ولی در گراف اصلی هست.

(زیر بخش چهارم)

خیر چرا که حلقه S-O-T-B یک حلقه چهارتایی است که بین هیچ دو راس غیر مجاورش هیچ بالی نیست.

(زیر بخش پنجم)

$$P(C, O, A, S, T, B, M) = \frac{1}{Z} \phi_1(C) \phi_2(O, S) \phi_3(B, S) \phi_4(O, T, A) \phi_5(A, T, M) \phi_6(B, T, M)$$

(زیر بخش ششم)

درست است. ●

$$P(O, A, S, T, B, M) = \int P(C) P(O) P(A) P(S|O) P(T|O, A) P(B|S) P(M|B, T, A) dc$$

$$P(O, A, S, T, B, M) = P(O) P(A) P(S|O) P(T|O, A) P(B|S) P(M|B, T, A)$$

چرا که هیچ عبارتی به جز $P(C)$ به C وابسته نیست و انتگرال برابر 1 است.

- درست است چرا که در حالت مارکوف توابع پتانسیل توزیع احتمال نیستند چرا که انتگرال الشان یک نیست و این امر با تقسیم بر Z جبران می شود و اگر انتگرال یک تابع پتانسیل یک باشد با حذف مقدار Z هم عوض نمی شود.

بخش سوم

(زیر بخش اول)

$$P(A, B, C, D, E, F, G) = \frac{1}{Z} \phi_1(A, B, C) \phi_2(B, C, D) \phi_3(C, D, F) \phi_4(F, E) \phi_5(B, E) \phi_6(E, G)$$

(زیر بخش دوم)

- a. نادرست. این دو راس در گراف بهم مسیر دارند پس مستقل نیستند.
- b. نادرست. با حذف D, C همچنان مسیر $F-E-B-A$ وجود دارد.
- c. درست. با حذف E راس G دیگر به هیچ کدام از رئوس دسترسی ندارد.
- d. درست است. چرا که B, C برابر با $Markov Blanket$ راس A هستند.

*(زیر بخش سوم)

با انجام این کار مقدار احتمال نرمالایز نشده 5 برابر می شود و به همین دلیل Z هم 5 برابر می شود تا اثر آن را خنثی کند پس توزیع احتمال (ضرب توابع پتانسیل نرمالایز شده) هیچ فرقی نخواهد کرد.

بخش چهارم

با توجه به شبکه بیزی که داریم:

$$P(X, Z) = P(Z) P(X|Z)$$

$$q(Z) = KL(q(z) || P(X, Z)) = E_q[\log \frac{q(z)}{P(Z, X)}] = E_q[\log \frac{\theta^2 z e^{-\theta z}}{z e^{-z(x+1)}}]$$

$$q(Z) = E_q[2\log\theta - \theta z + z(x+1)] = 2E_q[\log\theta] - E_q[\theta z] + E_q[z(x+1)]$$

از آنجایی که امید ریاضی روی q است پس به جز z بقیه متغیرها ثابت هستند. پس:

$$q(Z) = 2\log\theta - 2 * \theta * \frac{1}{\theta} + 2\frac{x+1}{\theta}$$

حالا برای یافتن θ بهینه مشتق می‌گیریم و مساوی صفر قرار می‌دهیم.

$$\frac{\partial q}{\partial \theta} = \frac{2}{\theta} - \frac{2x+2}{\theta^2} = 0 \Rightarrow \theta^* = x + 1$$

سؤال دوم

بخش اول

(زیر بخش اول)

در مدل VAE ما توزیع احتمال داده‌ها را در اختیار نداریم بلکه توزیع توأم را در اختیار داریم و محاسبه احتمال داده‌ها از لحاظ عملی نشدنی است. ترم دوم که فاصله فاصله توزیع فضای نهان از توزیع نرمال استاندارد را نشان می‌دهد و به دلیل خواص KL-Divergence مقدار کمینه‌اش صفر است همچنین فرض کردۀایم که z از توزیع نرمال با میانگین z می‌آید پس در توان این توزیع عبارت $D(z) = ||x||_2^2$ قرار دارد که این همان reconstruction loss است و سعی می‌کند که فاصله خروجی درست را با خروجی دیکودر کم کند پس اگر این عبارت کم شود مدل بهتر عمل می‌کند.

(زیر بخش دوم)

این دیتاست به عنوان یک محک برای مدل‌های مولدی که می‌خواهند disentanglement انجام دهند به کار می‌رود چرا که هر تصویر 6 بعد دارد و با تغییر مقادیر آن‌ها شکل‌ها عوض می‌شوند و شکل‌ها یا قلبند یا مستطیل یا بیضی یا دایره.

Random Sample of 8 Dsprites Images

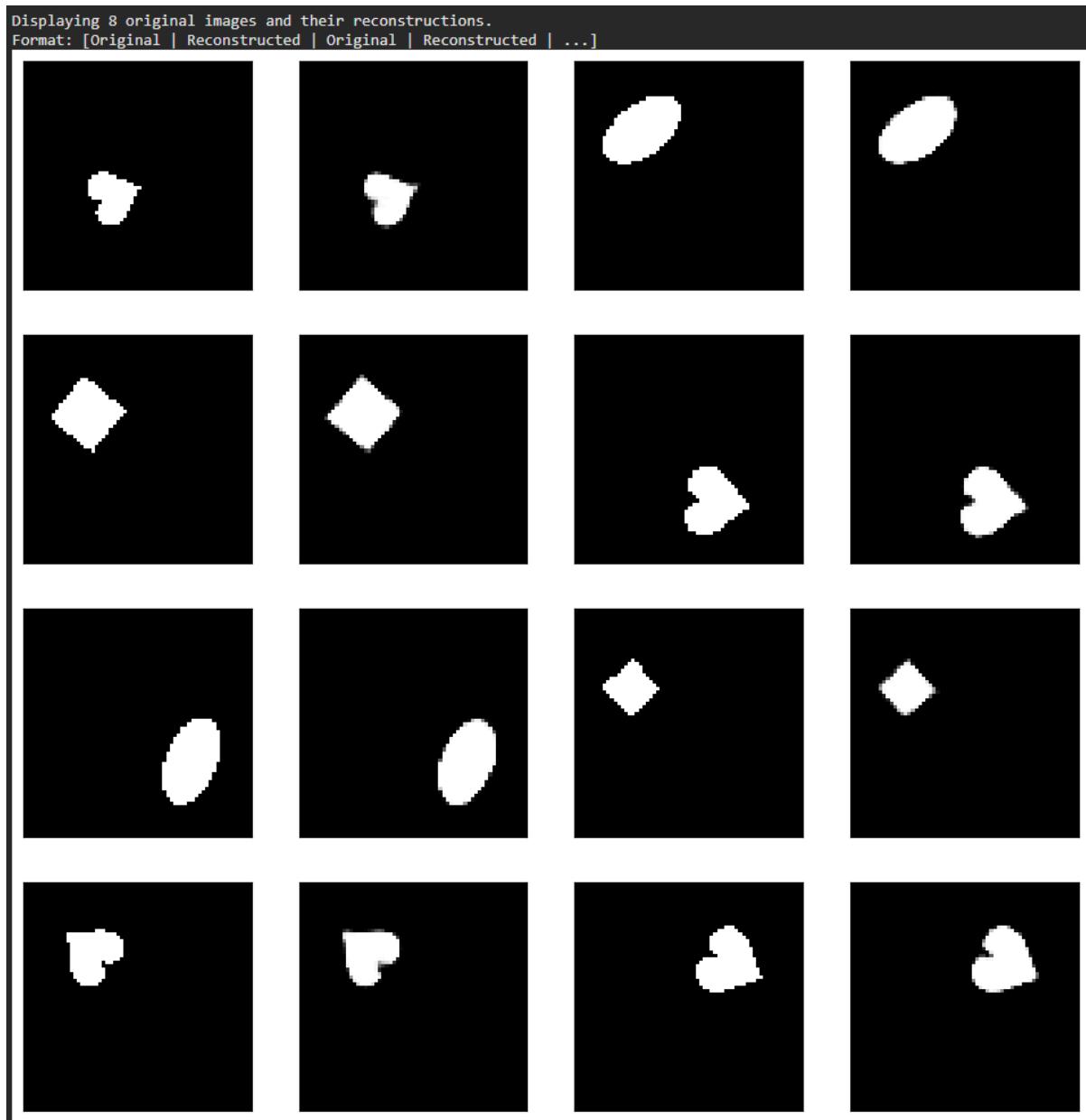


(زیر بخش سوم)

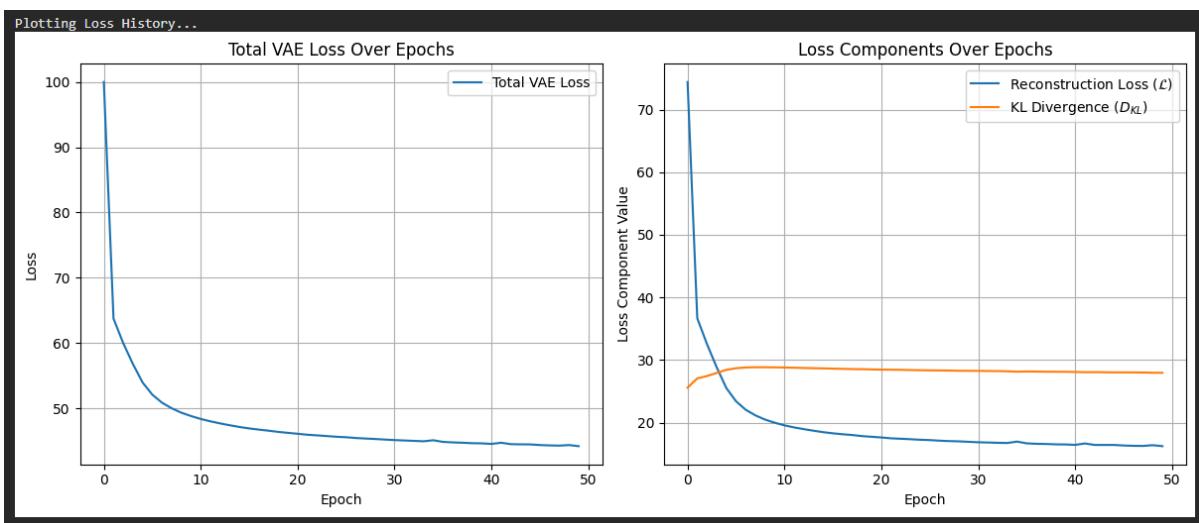
برای این کار از توزیع نرمال استاندارد یک نمونه تصادفی به نام ϵ می‌گیریم و آن را در انحراف معیار ضرب و با میانگین جمع می‌کنیم که هر دو هم خروجی‌های انکودر هستند اینگونه می‌توانیم مشتق بگیریم.

(زیر بخش چهارم)

نمونه‌ای از تصاویر تولید شده به وسیله این مدل



نمودار هزینه

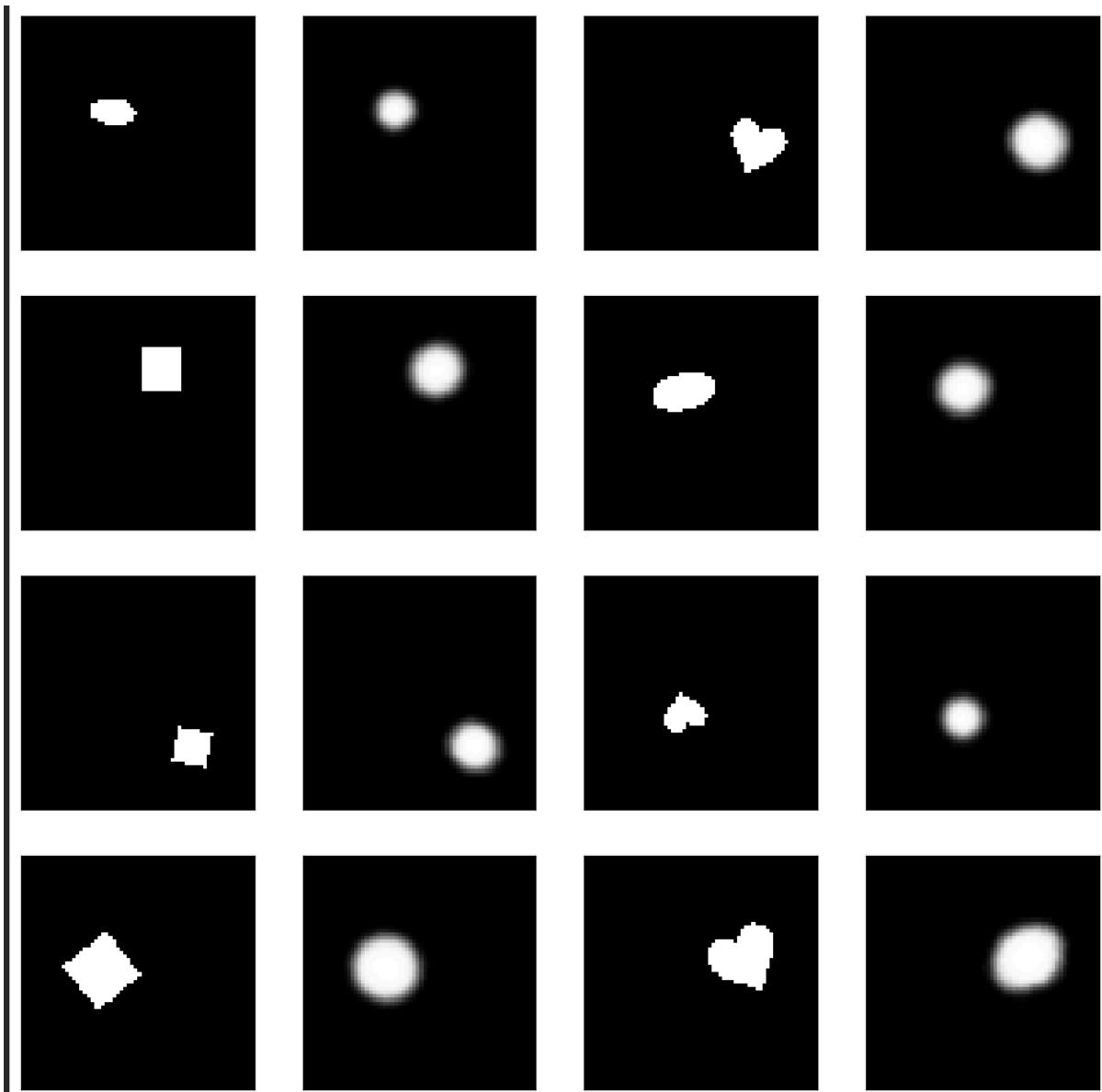


(زیر بخش پنجم)

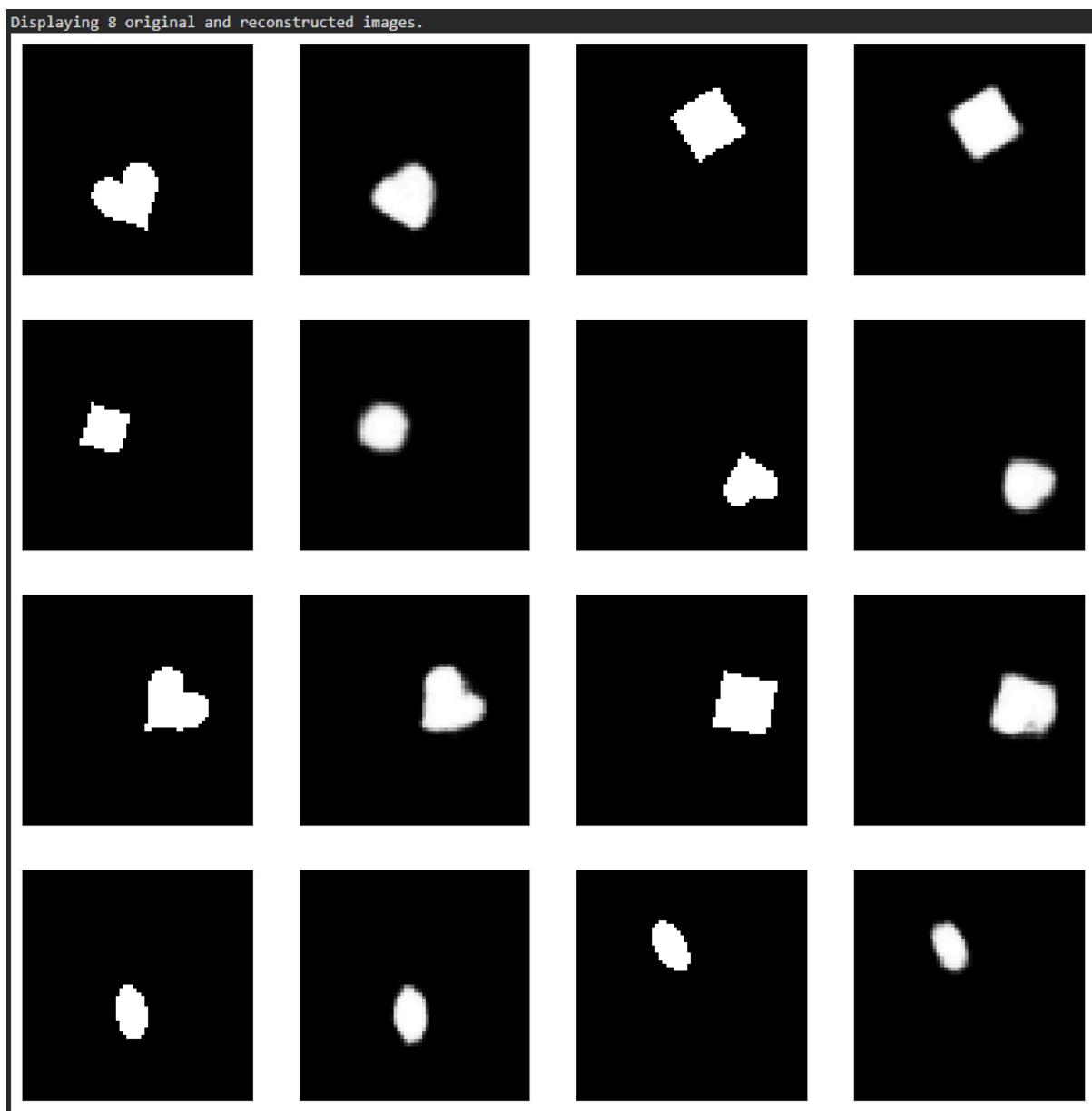
در این مدل سعی بر این است که ابعاد توزیع نهان معنا دار شوند و با اضافه کردن β در واقع بر روی نرمال شدن توزیع تأکید می کنم چرا که نرمال استاندارد نبودن آن پنالتی سنگینی خواهد داشت، حداقل سنگین تر از مدل عادی VAE.

(زیر بخش ششم)

بنتای 20



بتای 8



زیر بخش هفتم)

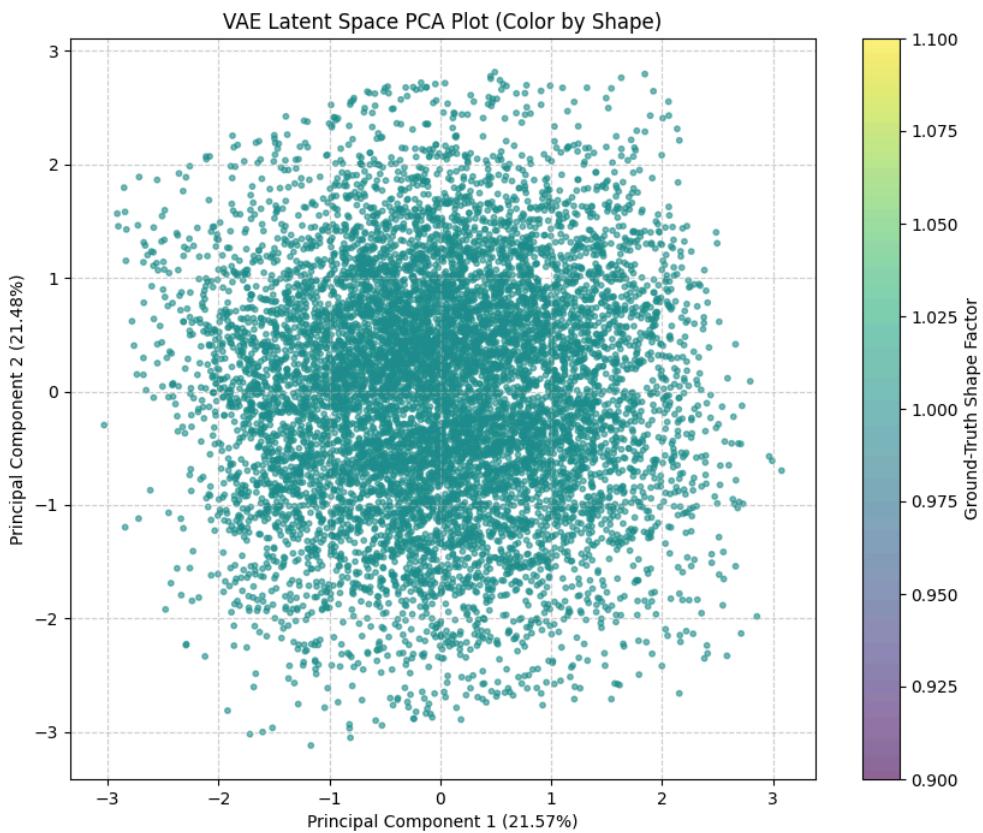
برای اندازه گیری استقلال ابعاد نهان از یکدیگر از این متريک استفاده می کنیم. معیار MIG ارزیابی می کند که یک بعد مشخص از فضای نهان (Z_j) تا چه اندازه بهطور انحصاری با یک عامل تغییر زمینه ای (Ground-Truth Factor) مانند شکل، مقیاس یا چرخش (τ_k) هم راستا شده است. این معیار برای تعیین درجه در همتبندگی به دست آمده توسط یک مدل مولد طراحی شده است.

بنا برابر با 1: 0.001

بنا برابر با 8: 0.1431

بنا برابر با 20: 0.3764

زیر بخش هشتم)



همانگونه که در تصویر برای بتای 8 مشاهده می شود نمونه ها تقریباً شبیه به توزیع نرمال استاندارد در وسط پخش شده‌اند و همانگونه که گفته شد MIG این مدل پایین است و هر بعد تنها حدود 20 درصد داده ها را دارد و این یعنی هیچ بعد غالبی وجود ندارد پس disentanglement خیلی خوبی حاصل نشده است.

بخش دوم

(زیر بخش اول)

تفاوت اصلی VQ-VAE با VAE استاندارد این است که فضای نهان گستته (Discrete Latent Space) یک فضای نهان VQ-VAE است که استاندارد از فضای نهان پیوسته (مانند توزیع گوسی) استفاده می‌نماید. این فرآیند، نقش نگاشت خروجی پیوسته رمزگذار (x_e) را به نزدیکترین بردار (z_q) در یک مجموعه متناهی و قابل یادگیری از بردارها ایفا می‌کند که به آن **دفترچه کد** (Codebook) یا فضای امبدینگ گفته می‌شود.

مزایای این روش توزیع نهان ساده‌تر و نگاشت راحت‌تر خواص به ابعاد مختلف است.

(زیر بخش دوم)

این روش پیشنهاد می‌کند که به جای توزیع نرمال استاندارد از یک mixture of Gaussians استفاده کنیم تا مدل انعطاف و قدرت مدل سازی بیشتری بگیرد پس توزیع پیشین تبدیل می‌شود به:

$$P(z) = \frac{1}{K} \sum_{k=1}^K P(z | u_k)$$

که u_k ورودی‌های کاذب قابل یادگیری هستند و مولفه‌های توزیع ما حول آن‌ها متumerکزند.

توزیع پیشین نرمال استاندارد اغلب برای تطبیق با توزیع پیچیده متغیرهای نهان رمزگذاری شده بسیار ساده است.

این عدم تطابق باعث می‌شود که جمله واگرایی KL اغلب غالب شود و نمایش‌های رمزگذاری شده را به سمت فروپاشی در یک منطقه کوچک از فضای نهان سوق دهد.

با استفاده از **VampPrior**، مدل به یک توزیع پیشین وابسته به داده دست می‌یابد که انعطاف‌پذیری بیشتری داشته و بهتر می‌تواند با پسین تجمعی واقعی مطابقت داشته باشد.

این تطابق بهتر محدودیت جمله واگرایی KL را کاهش می‌دهد و به رمزگذار اجازه می‌دهد تا فضای نهان را بهطور موثر تر و گسترده‌تر مورد استفاده قرار دهد، که در نهایت منجر به بیبود کیفیت نمونه‌های تولیدی نسبت به توزیع پیشین گوسی استاندارد می‌شود.

زیر بخش سوم

SC-VAE با هدف معرفی مفهوم کدگذاری تنک (Sparse Coding - SC) در چارچوب VAE طراحی شده است. کدگذاری تنک که داده‌ها می‌توانند به عنوان ترکیبی خطی از تعداد کمی از بردارهای پایه نمایش داده شوند که ضرایب این ترکیب اکثراً صفر هستند (تنک).

الگوریتم آستانه‌گذاری انقباضی تکراری (ISTA) یک الگوریتم متدالول برای حل مسئله کدگذاری تنک (یافتن ضرایب تنک) است.

در **SC-VAE**، نویسنده‌گان شبکه رمزگذار را به‌گونه‌ای طراحی می‌کنند که مراحل تکراری ISTA را تقلید کند. بهطور خاص، شبکه استنتاج به عنوان یک شبکه **L-ISTA** (ISTA یادگرفته شده) ساخته می‌شود.

این طراحی، رمزگذار را مجبور می‌کند تا نمایش‌های نهان تولید کند که ذاتاً تنک هستند.

نمایش‌های تنک اغلب قابل تفسیر تر هستند، زیرا تنها تعداد کمی از ویژگی‌ها برای هر ورودی فعل می‌شوند، که درک اینکه کدام عوامل نهان، بازسازی را هدایت می‌کنند، آسان‌تر می‌سازد.

کدهای تنک اغلب در برابر نویز و ویژگی‌های نامربوط در داده‌های ورودی قوی‌تر هستند.

مراجع