

پروژه درس یادگیری ماشین ترم ۴۰۴۱

اعضای گروه: حامد امینی - علی حیدری - علی کریمی.

شماره پروژه: ۵

عنوان پروژه: **Fashion MNIST VAE**

آدرس مخزن کد پروژه: <https://github.com/HamedAV/ML Project Fashion-MNIST-VAE>

فاز اول: آماده سازی داده ها و تحلیل اکتشافی

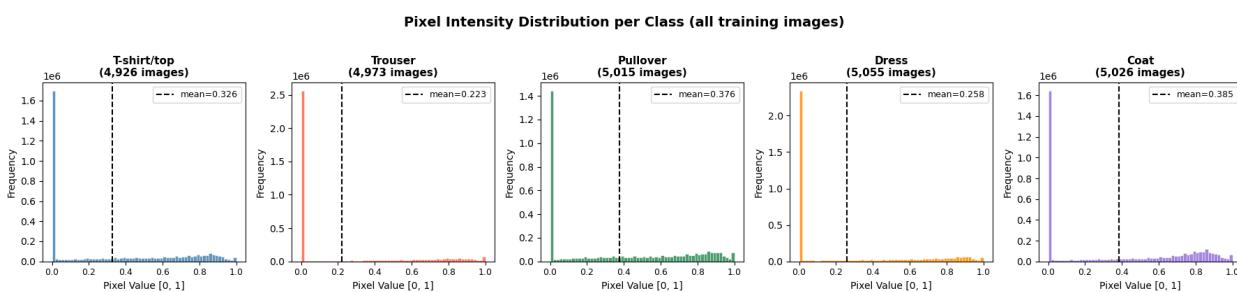
خروجی sanity check برای کلاس های مختلف به این شکل است:



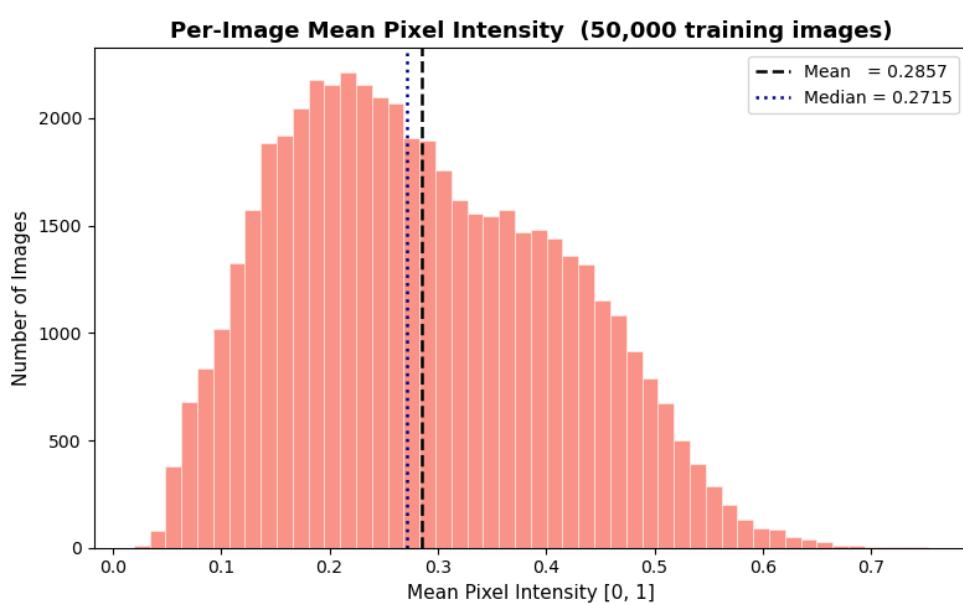
نمودار میله ای توزیع کلاس ها



هر کلاس تقریباً ۵۰۰۰ نمونه دارد
خروجی هیستوگرام شدت پیکسل‌ها:



جمع پیکسل‌ها در ۰ برای قسمت‌های پس زمینه مشکی عکس‌ها است.
تحلیل پیکسل‌های غیر صفر:

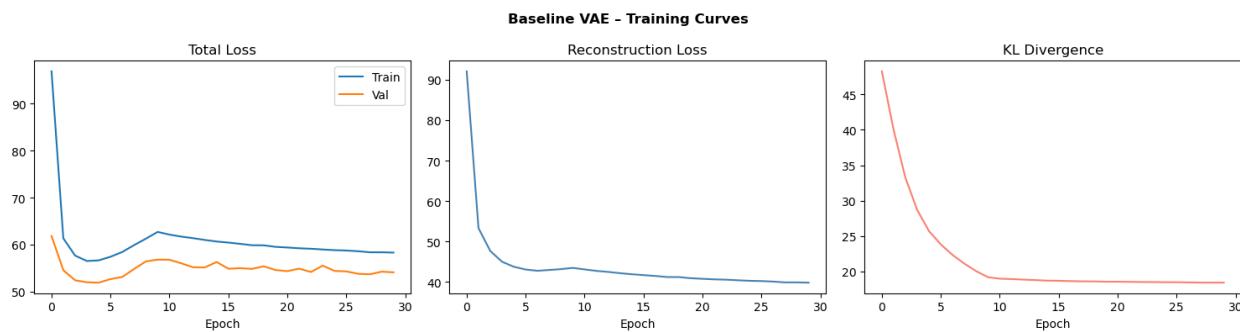


فاز دوم: پیاده سازی

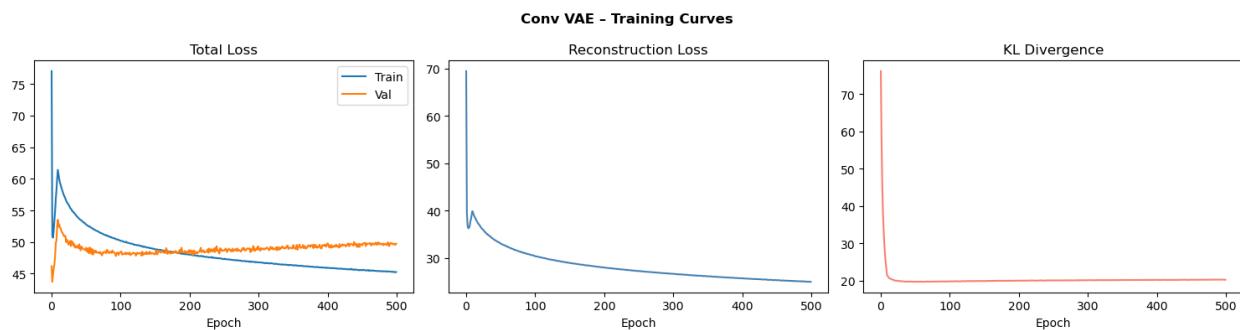
پایه با معماری VAE latent_dim = 32 و FC (fully-connected) Loss پیاده سازی شد بر اساس فرمول زیر محاسبه شد:

$$\text{Total Loss} = \text{Reconstruction Loss} + \text{Regularization Factor}$$

$$\text{Regularization Factor} = \beta \times \text{KL Divergence}$$



مدل بهبودیافته با معماری latent_dim = 64 ، Convolutional (ConvVAE) کیفیت بهتر توسعه یافت.



جدول مقایسه

Model / Configuration	Total Loss	Reconstruction Loss	KL Divergence
Baseline VAE (FC, $\beta = 1$)	54.3819	35.7678	18.6140
Improved VAE (Conv, $\beta = 1$)	49.9575	29.9687	19.9887

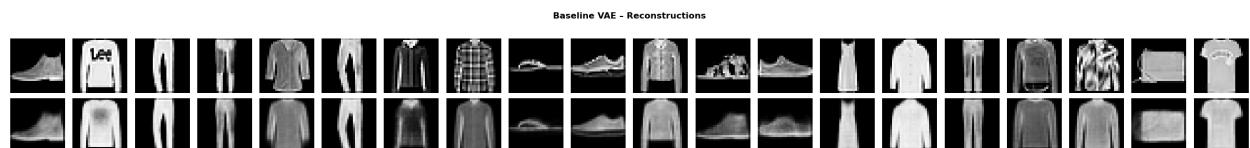
جدول نهایی نتایج روی مجموعه داده تست

Model / Configuration	Total Loss	Reconstruction Loss	KL Divergence
Baseline VAE (FC, $\beta = 1$)	54.3819	35.7678	18.6140
Improved VAE (Conv, $\beta = 1$)	49.9575	29.9687	19.9887
Conv VAE ($\beta = 0.5$)	40.6997	25.6985	30.0024
Conv VAE ($\beta = 1$)	50.2298	30.1287	20.1011
Conv VAE ($\beta = 4$)	84.7149	41.7747	10.7351
cVAE (Conv, $\beta = 1$)	51.82	34.74	17.08

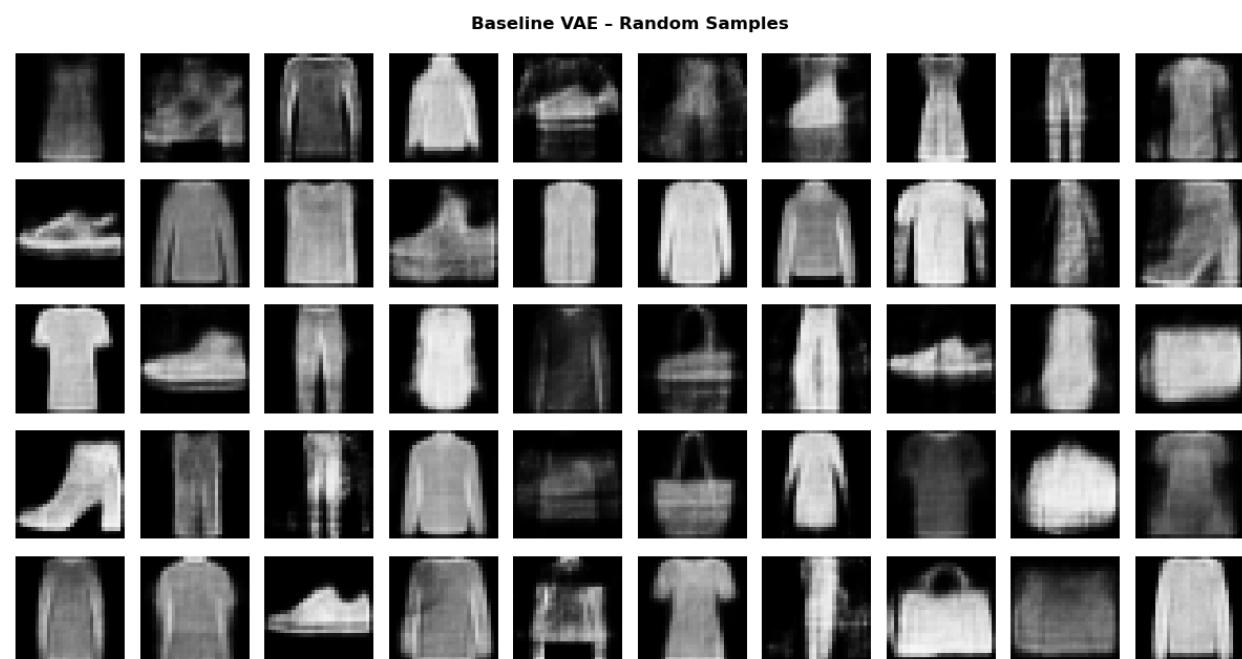
FID Score (Conv VAE): **10.5457**

نمایش نمونه های کلیدی بازسازی (Sampling) و نمونه برداری (Reconstruction) (Baseline VAE)

بازسازی

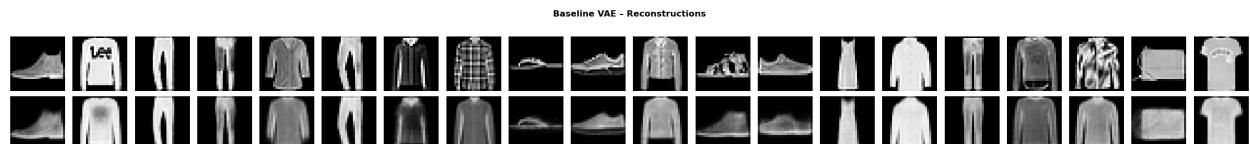


نمونه برداری

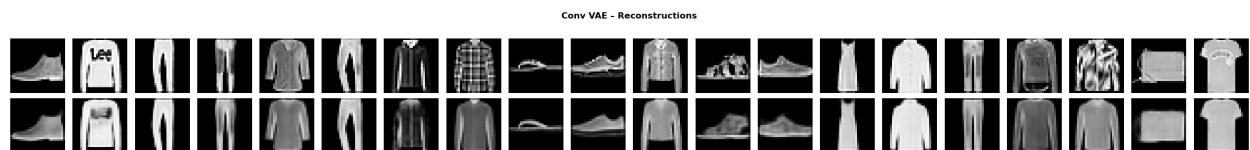


مقایسه قبل و بعد برای بهبود کیفیت خروجی (Baseline vs Improved VAE)

بازسازی قبل از بهبود



بازسازی بهبود یافته

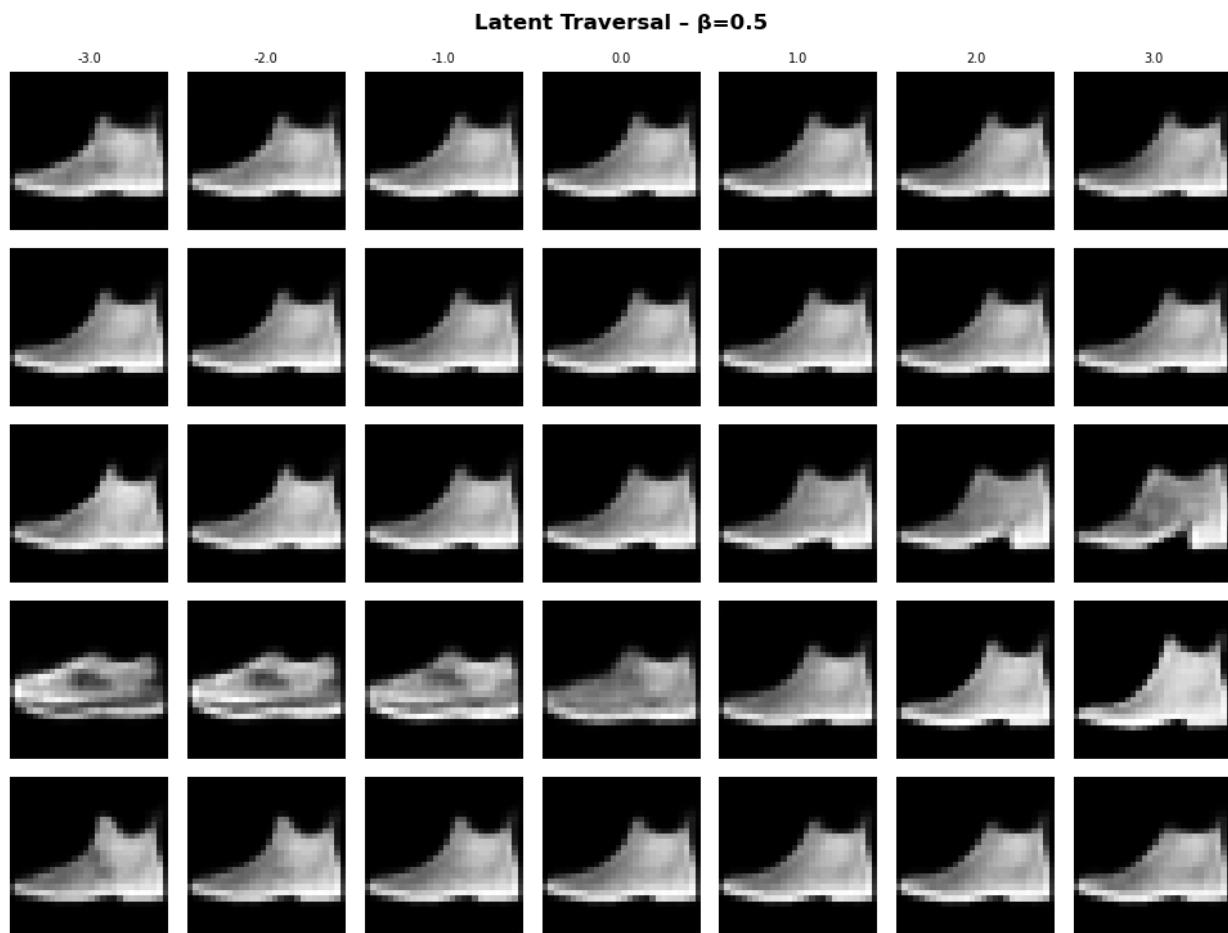


مدل بهبود یافته (Conv) کیفیت بهتری دارد (کمتر بلور، جزئیات بیشتر) اما ممکن است KL بالاتری داشته باشد.

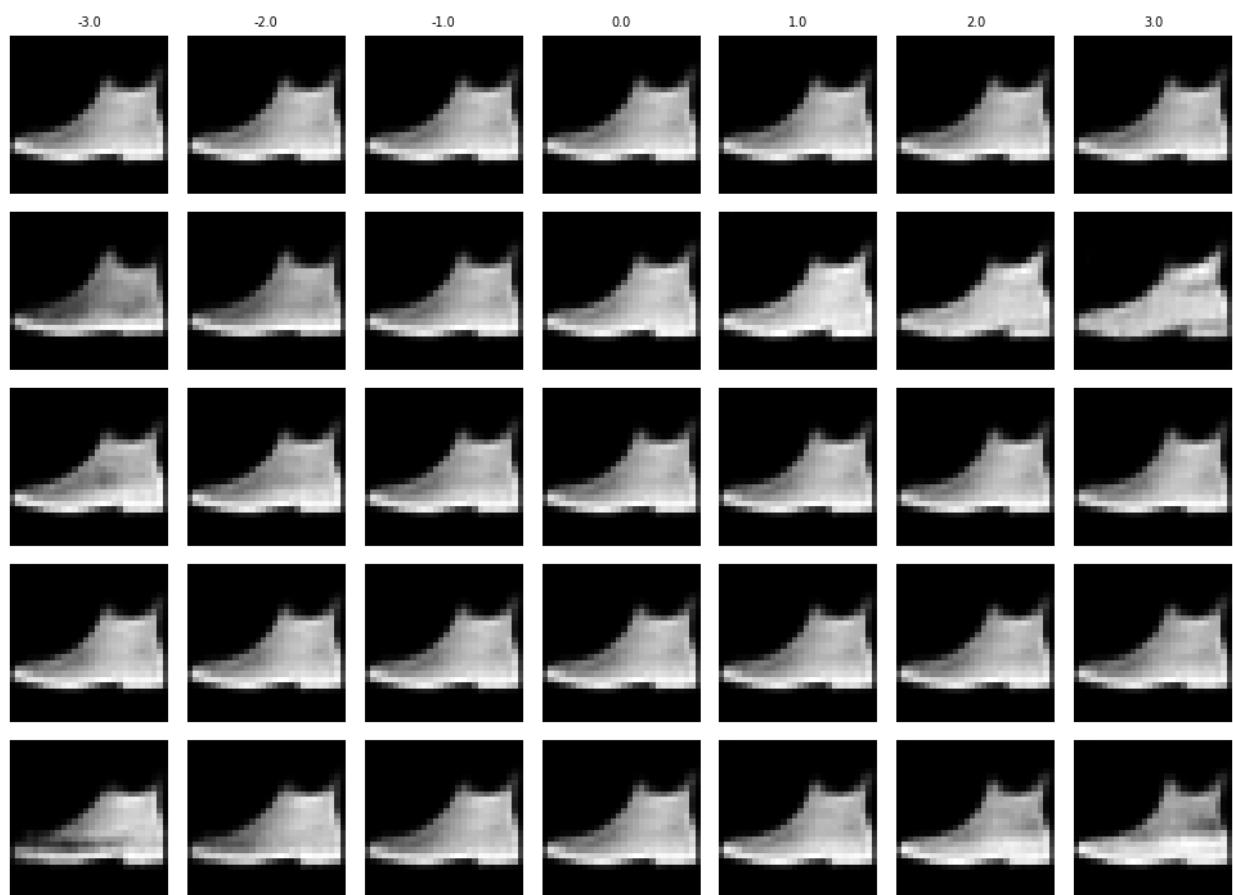
Baseline VAE Conv VAE



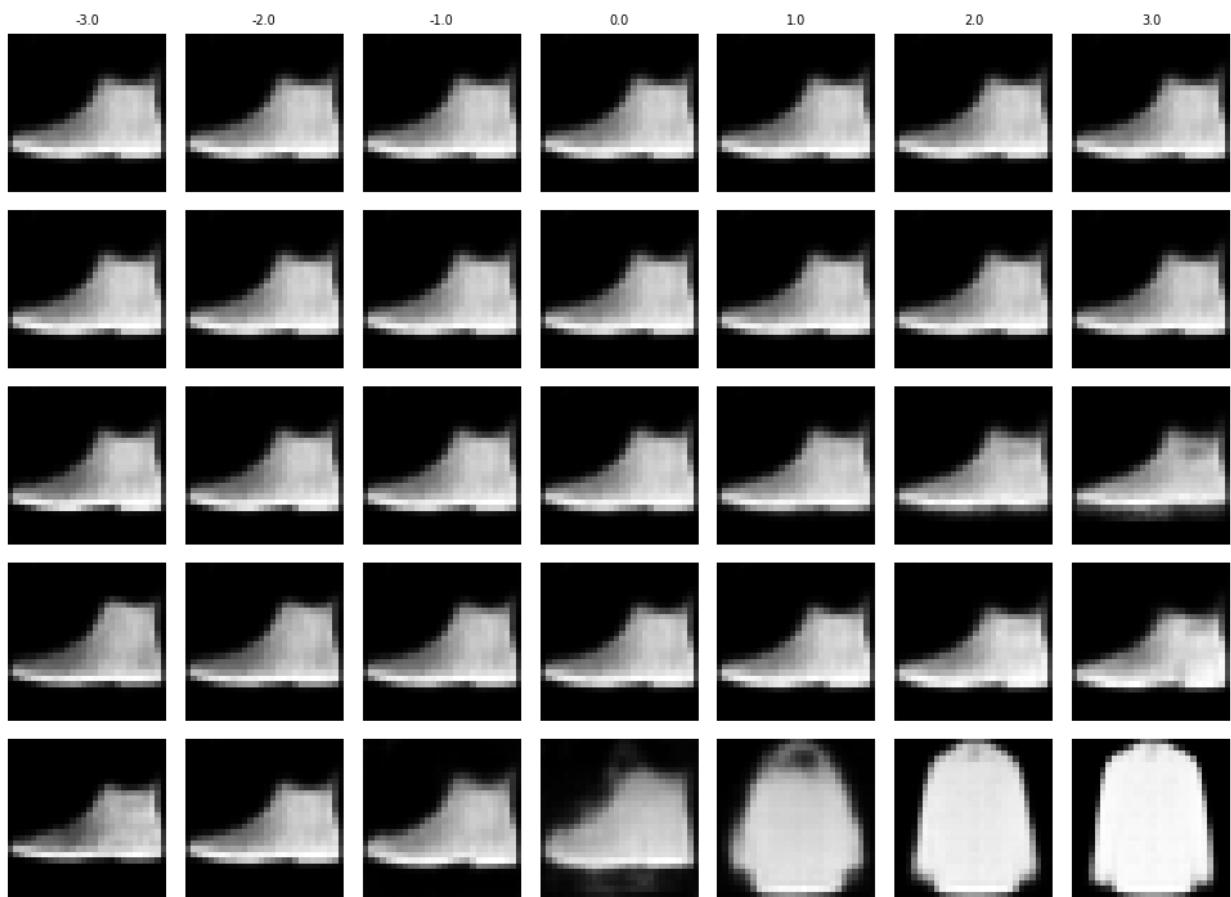
پیمایش فضای نهفته (Latent Traversal) برای وزن‌های مختلف (β)

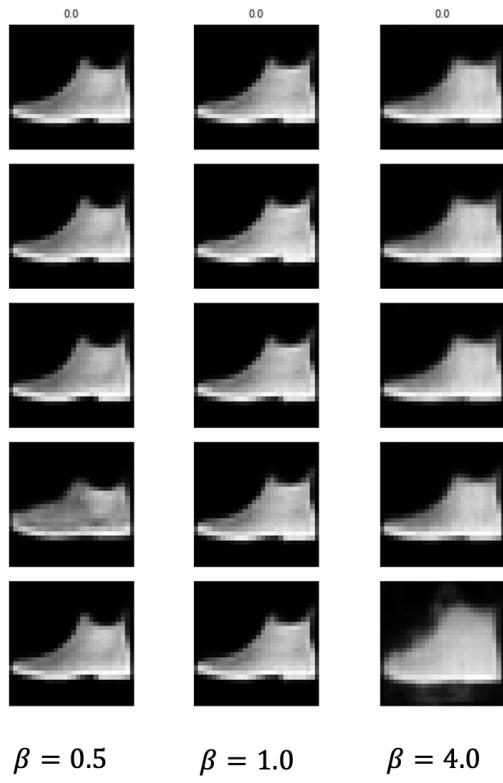


Latent Traversal - $\beta=1.0$



Latent Traversal - $\beta=4.0$

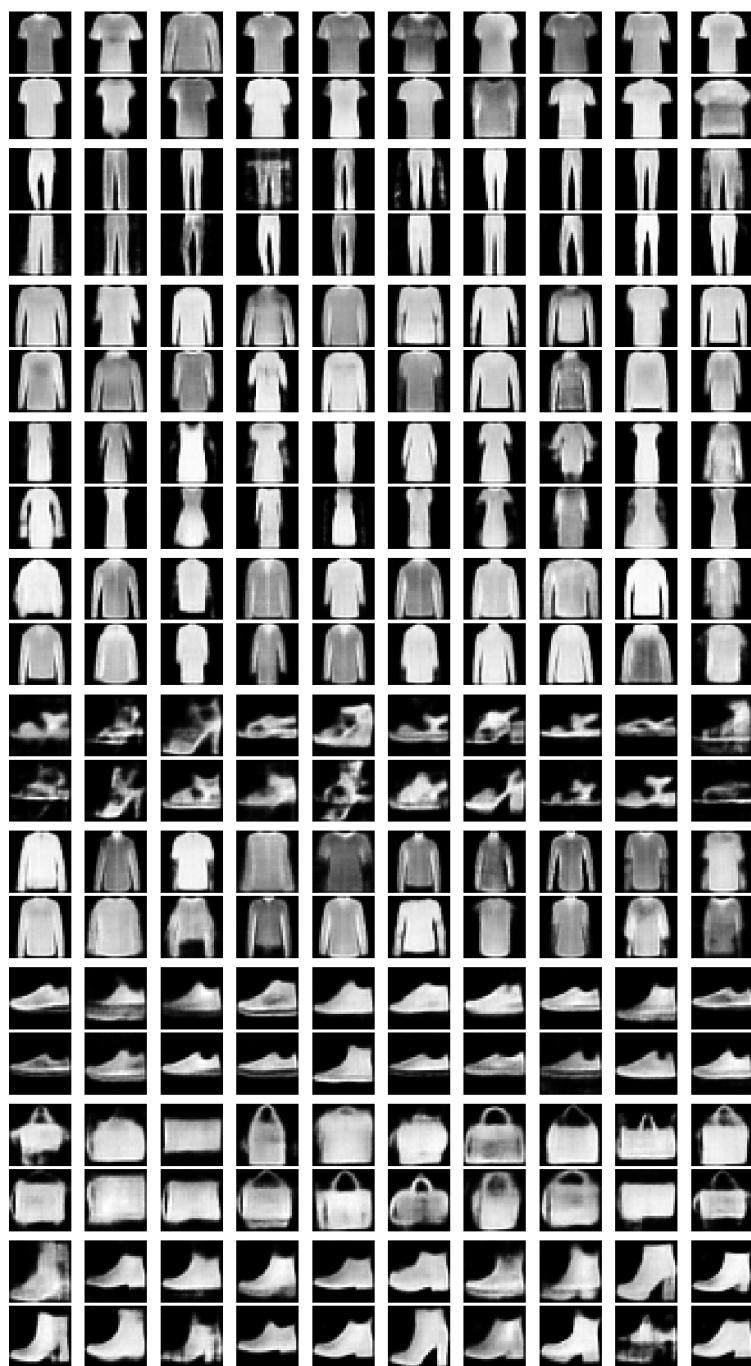




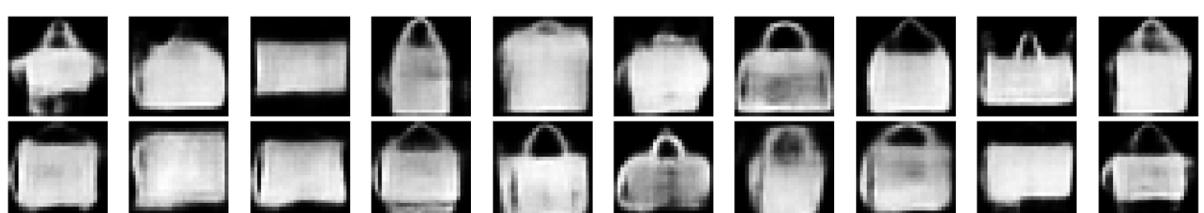
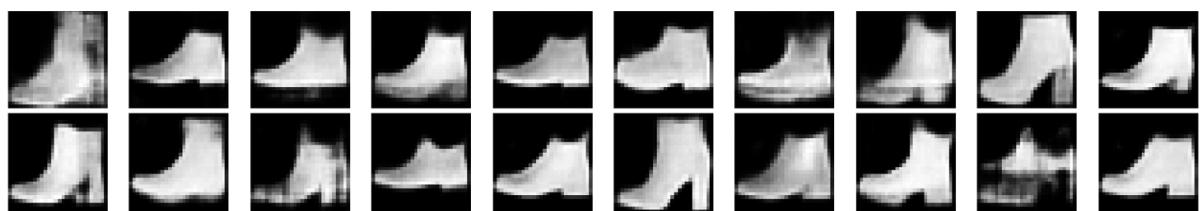
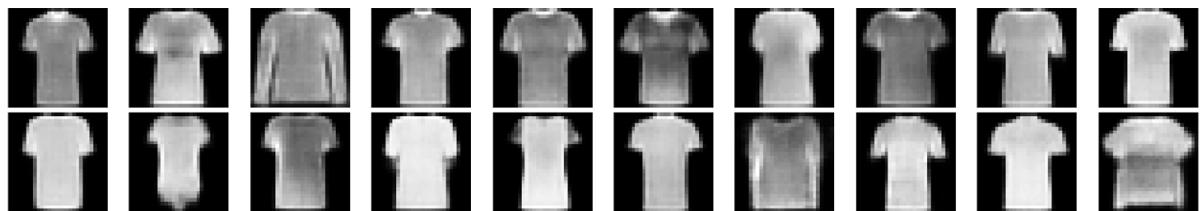
با افزایش β ، فضای نهفته منظم‌تر و disentangled تر می‌شود (تغییرات معنادارتر در ابعاد)، اما کیفیت بازسازی کاهش می‌یابد.

تولید کنترل‌پذیر با برچسب برای تمام کلاس‌ها (cVAE)

cVAE - Class-Conditional Generation (20 per class)

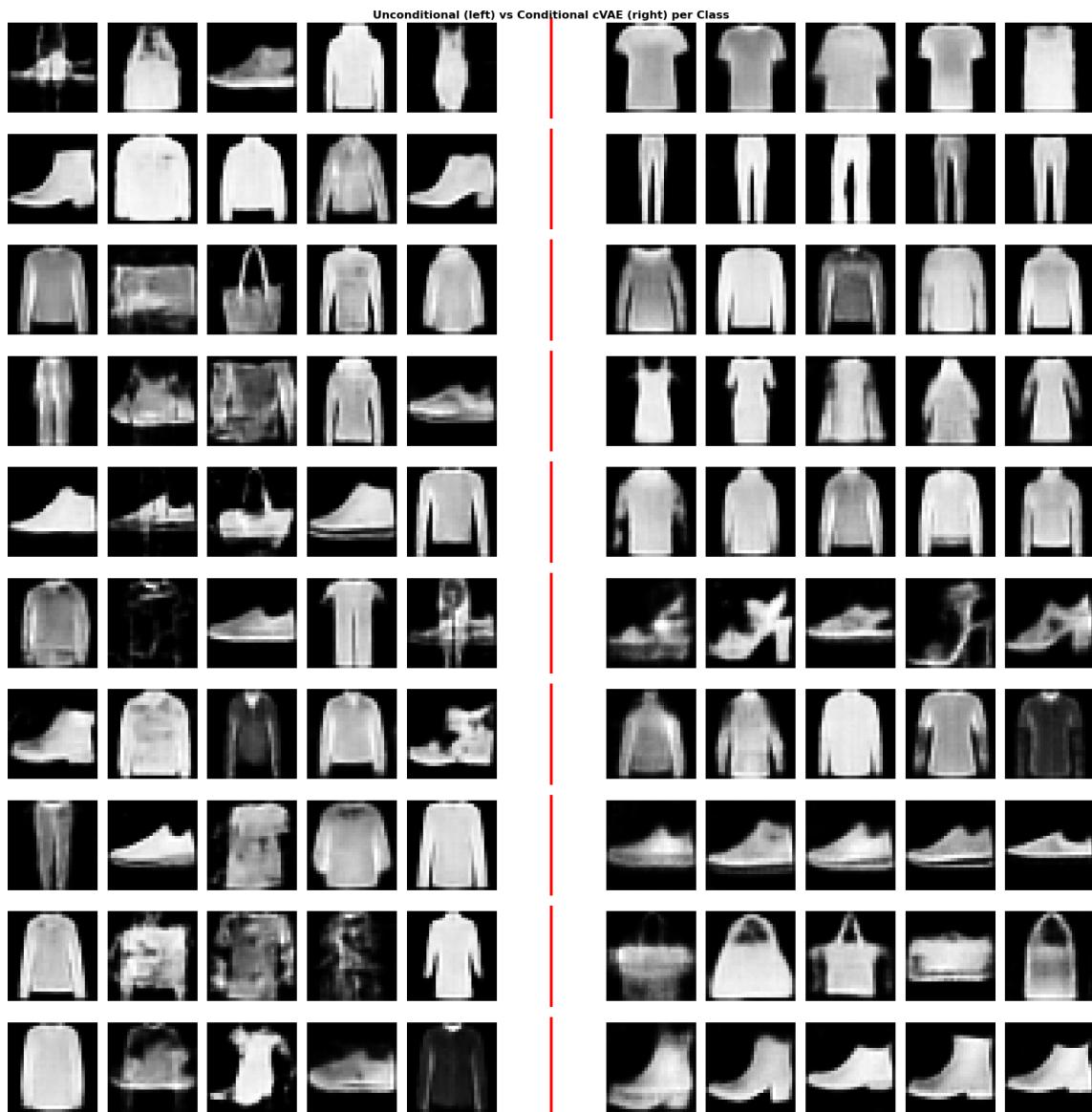


بررسی ۳ کلاس خاص:



مقایسه با VAE غیرشرطی

cVAE تولیدات کلاس محور دارد، در حالی که VAE معمولی مخلوط است.



```

==== cVAE Controllability Evaluation ====
Evaluating controllability: 0% | 0/5000 [00:00<?, ?it/s]
    T-shirt/top (class 0): 90.2%
    Trouser (class 1): 99.0%
    Pullover (class 2): 83.4%
    Dress (class 3): 90.2%
    Coat (class 4): 25.4%
    Sandal (class 5): 95.0%
    Shirt (class 6): 72.8%
    Sneaker (class 7): 88.8%
    Bag (class 8): 94.4%
    Ankle boot (class 9): 96.0%

Mean Accuracy: 83.52%

```

نتیجه گیری

بهترین تنظیمات: بر اساس جدول نتایج و نمونه‌ها، مدل Improved Conv VAE با $\beta = 1$ بهترین تعادل را بین کیفیت تولید (Recon FID مناسب) و منظم‌سازی (KL متعادل) ارائه می‌دهد. برای تولید کنترل‌پذیر، cVAE برتر است (دقت طبقه‌بندی بالا، تنوع درون‌کلاسی خوب).

β بالاتر (مانند 4) disentanglement بهتری در فضای نهفته می‌دهد اما خروجی‌ها بلور تر هستند.

محدودیت‌ها

- مدل‌ها روی تصاویر خاکستری ساده 28×28 خوب کار می‌کنند اما ممکن است برای داده‌های پیچیده‌تر (رنگی، بزرگ‌تر) نیاز به معماری عمیق‌تر داشته باشند.
- FID فقط برای یک مدل محاسبه شده؛ ممکن است overfit به طبقه‌بند خاص باشد.
- cVAE به برچسب‌ها وابسته است؛ در داده‌های بدون برچسب کارایی کمتری دارد.

پیشنهادات توسعه:

- آزمایش با latent_dim بالاتر (مثلاً 128) برای بهبود کیفیت.
- افزودن تکنیک‌هایی مانند Flow-based disentanglement یا VQ-VAE برای بهتر.
- ارزیابی روی داده‌های واقعی (نه MNIST-like) و محاسبه متريک‌های بيشتر مانند IS Score.
- ادغام با GAN برای خروجی‌های شارپ‌تر (VAE-GAN).