

实验报告：深度卷积生成对抗网络（DCGAN）

姓名：廖望

学号：2210556

[代码仓库](#)

1. 实验概述与核心原理

1.1. 实验目标

本实验的核心目标是实现并训练一个**深度卷积生成对抗网络 (Deep Convolutional Generative Adversarial Network, DCGAN)**，使其能够学习并模拟 **FashionMNIST** 数据集的内在分布。我们旨在通过模型训练，让**生成器 (Generator)** 网络能够从一个随机的潜在向量 (latent vector) 中，创造出足以以假乱真、在视觉上与真实衣物图像无法区分的、高质量的伪造样本。

1.2. GAN 的"零和博弈"思想

生成对抗网络 (GAN) 的本质是一场精妙的**双人零和博弈**。博弈的双方是：

- 生成器 (G)**：一个伪造大师，其目标是学习真实数据的分布，创造出尽可能逼真的伪造数据来"欺骗"判别器。
- 判别器 (D)**：一个鉴定专家，其目标是尽可能准确地分辨出输入数据是来自真实数据集，还是由生成器伪造的。

二者在训练中不断对抗、共同进化：D 通过学习变得越来越"火眼金睛"，而 G 为了骗过 D，也不得不生成越来越逼真的数据。这个过程最终的理想状态是达到**纳什均衡 (Nash Equilibrium)**，此时 G 生成的数据与真实数据无异，D 只能靠随机猜测来判断真伪（即判断概率为0.5）。

1.3. DCGAN 架构：稳定 GAN 训练的关键

本实验采用的DCGAN架构，是原始GAN的一个重要里程碑式改进，它通过一系列架构约束，极大地提升了GAN训练的稳定性：

- 用卷积层替代全连接层**：在判别器中使用带步长 (Strided) 的卷积层，在生成器中使用转置卷积层，让网络能学习到图像的空间层级特征。

- **消除池化层**：完全摒弃最大池化等下采样层，让网络自行学习空间下采样/上采样的方式。
- **引入批归一化 (BatchNorm)**：在G和D中都使用批归一化，以稳定数据流，缓解模式崩溃 (mode collapse)。
- **使用 LeakyReLU 和 Tanh**：在判别器中使用 LeakyReLU 作为激活函数以防止梯度稀疏，在生成器的最后一层使用 Tanh 激活函数。

1.3.1. 生成器 (Generator) 结构

生成器负责将一个100维的潜在向量，通过一系列的**转置卷积 (ConvTranspose2d)** 和批归一化操作，逐步上采样，最终映射为一张 1x28x28 的图像。

```
Generator(  
    (main): Sequential(  
      (0): ConvTranspose2d(100, 512, 4, 1, 0, bias=False)  
      (1): BatchNorm2d(512, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)  
      (2): ReLU()  
      (3): ConvTranspose2d(512, 256, 3, 2, 1, bias=False)  
      (4): BatchNorm2d(256, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)  
      (5): ReLU()  
      (6): ConvTranspose2d(256, 128, 4, 2, 1, bias=False)  
      (7): BatchNorm2d(128, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)  
      (8): ReLU()  
      (9): ConvTranspose2d(128, 1, 4, 2, 1, bias=False)  
      (10): Tanh()  
    )  
)
```

注：为简洁起见，此处省略了重复的模块参数。

1.3.2. 判别器 (Discriminator) 结构

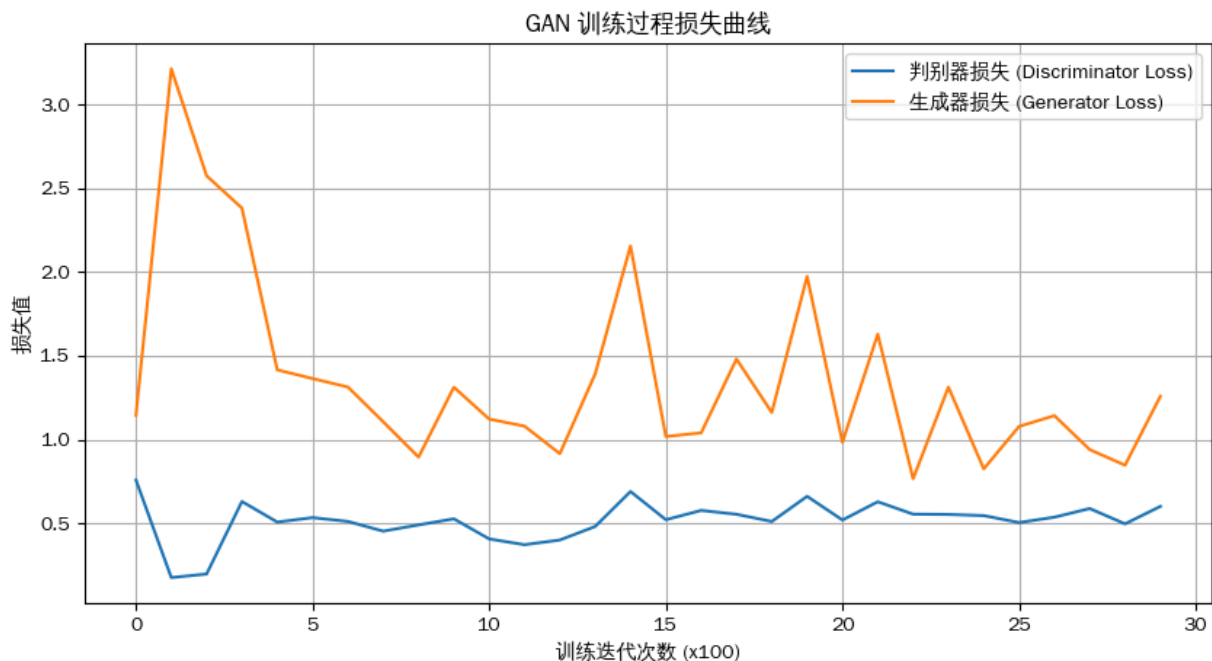
判别器则执行逆向操作，通过一系列带步长的**卷积 (Conv2d)** 和 LeakyReLU 激活，将输入的 1x28x28 图像下采样，最终输出一个0到1之间的标量，代表该图像为"真"的概率。

```
Discriminator(  
    (main): Sequential(  
      (0): Conv2d(1, 64, 4, 2, 1, bias=False)  
      (1): LeakyReLU(negative_slope=0.2)  
      (2): Conv2d(64, 128, 4, 2, 1, bias=False)  
      (3): BatchNorm2d(128, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)  
      (4): LeakyReLU(negative_slope=0.2)  
      (5): Conv2d(128, 256, 3, 2, 1, bias=False)  
      (6): BatchNorm2d(256, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)  
      (7): LeakyReLU(negative_slope=0.2)  
      (8): Conv2d(256, 1, 4, 1, 0, bias=False)  
      (9): Sigmoid()  
    )  
)
```

注：为简洁起见，此处省略了重复的模块参数。

2. 训练动态：一场永不停歇的"军备竞赛"

下图展示了在 FashionMNIST 数据集上训练15个轮次（Epochs）过程中，生成器和判别器损失值的变化情况。



GAN 的训练动态与传统监督学习截然不同，它是一场激烈的博弈过程，其损失曲线揭示了这场“军备竞赛”的本质：

- **判别器损失 (D Loss):** 判别器的损失（蓝线）在整个训练过程中始终保持在较低的水平（约0.5-1.0之间）且相对稳定。这表明判别器是一个高效的学习者，能持续且敏锐地分辨出真实图像与生成器当时的伪造品。一个强大的判别器是GAN成功训练的基石，因为它能为生成器提供准确且有意义的梯度信号。
- **生成器损失 (G Loss):** 生成器的损失（橙线）在训练中表现出剧烈的、高频的波动。这并非训练不稳定的信号，恰恰相反，它反映了**健康的对抗性学习过程**。当G的损失突然飙升时，通常意味着D刚刚“识破”了G的某种伪造策略，导致G的“欺骗”行为大量失败。而G为了应对，必须快速学习新的、更高级的伪造技巧，从而使其损失再次下降。这种震荡是两者在向着纳什均衡动态逼近时的必然表现。
- **动态平衡:** 尽管G的损失波动剧烈，但两个模型的损失并未出现单向发散（例如一方损失趋近于0而另一方无限增大），而是维持在一定的范围内振荡。这表明训练过程在宏观上是收敛的，形成了一种动态平衡，双方的能力都在螺旋式上升。

3. 生成结果与潜在空间分析

3.1. 随机样本生成质量

我们从训练好的生成器中随机采样8个潜在向量并生成了对应图像。这些图像质量较高，轮廓清晰，展示了模型学习到的多样性，涵盖了T恤、裤子、套衫、连衣裙等多种衣物类别，证明模型没有发生严重的模式崩溃。



3.2. 潜在空间解耦特性探索

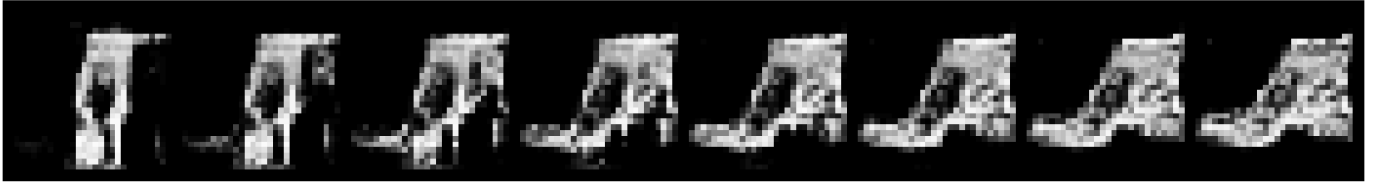
为了探究模型是否学习到了一个有意义的、**解耦 (disentangled)** 的潜在空间，我们进行了一系列维度操控实验。我们固定一个基准向量，然后微调其中某一个维度的值（从-4.0到+4.0），观察生成图像的连续变化。一个理想的解耦表示意味着单个潜在维度应控制着某个单一的、可解释的视觉特征。

3.2.1. 宏观语义维度 (Controlling Macro-Semantics)

在某些基准向量上，我们发现有些维度清晰地控制着高级的、语义层面的特征。

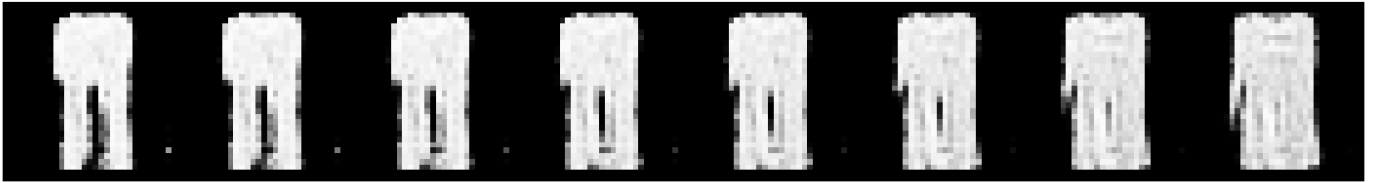
- **基准向量5 (鞋履):** 维度5控制着鞋子的子类别。当值从负到正变化时，图像清晰地从一个**高帮靴/运动鞋**平滑地过渡到一个**凉鞋或开口鞋**的形态。

基准向量 5, 调整维度 5



- **基准向量99 (裤子):** 维度5清晰地控制了**裤子的宽度**。图像从左到右，裤腿由宽松的阔腿裤样式逐渐收窄为紧身的锥形裤。

基准向量 99, 调整维度 5

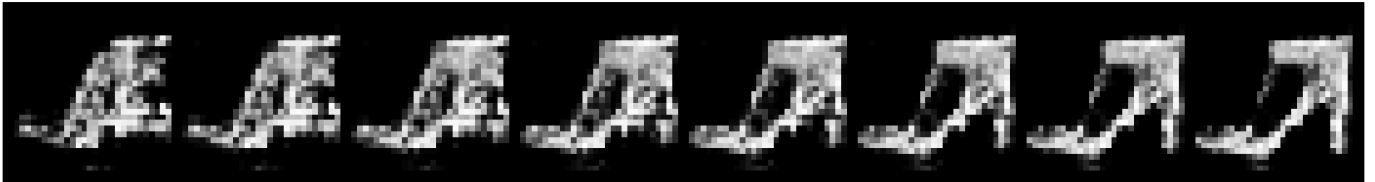


3.2.2. 微观属性维度 (Controlling Micro-Attributes)

另一些维度则控制着更细微的、非类别相关的视觉属性。

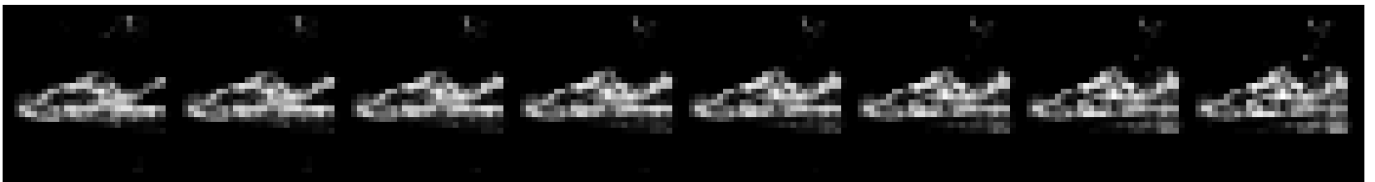
- **基准向量5 (鞋履):** 维度15主要控制**鞋子的观察视角**。随着值的变化，鞋子发生了轻微的旋转，展示出更多侧面或正面的轮廓。

基准向量 5, 调整维度 15



- **基准向量42 (运动鞋):** 维度25影响了**鞋子侧面的图案或纹理**。可以看到随着值的变化，鞋侧的细节（如条纹）会发生改变。

基准向量 42, 调整维度 25

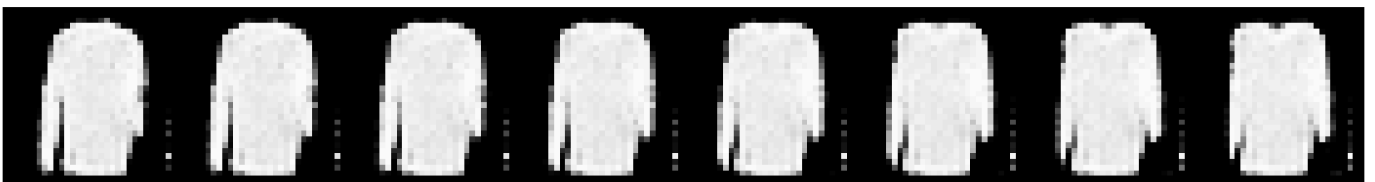


3.2.3. 非语义与纠缠维度 (Non-Semantic & Entangled Dimensions)

我们也观察到，并非所有维度都有清晰的语义含义，这反映了完全的特征解耦是极为困难的。

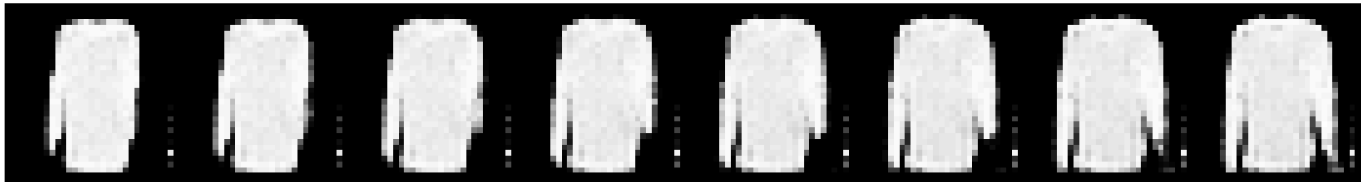
- **基准向量33 (套头衫):** 维度15控制了**图像的整体质量**，如清晰度与噪点水平，而非衣物本身的属性。

基准向量 33, 调整维度 15



- **基准向量33 (套头衫):** 维度25是一个**结构破坏维度**。调整它不会改变衣物的款式，而是会引入随机的伪影，导致图像结构逐渐崩坏。

基准向量 33, 调整维度 25



- **基准向量18 (套头衫):** 在此基准上，多个维度的影响都非常细微或相互**纠缠**，难以归纳出其控制的单一视觉特征。这表明潜在空间的某些方向可能是冗余的或编码了混合信息。

4. 实验总结与展望

本次实验通过成功构建并训练一个DCGAN模型，深刻揭示了生成对抗网络在无监督学习和高维数据分布建模方面的强大能力。

- **博弈式训练的有效性:** 我们成功复现了生成器与判别器之间的"军备竞赛"，并从损失曲线中解读出其健康的动态平衡，验证了GAN作为一种学习范式的有效性。
- **潜在空间的结构化特性:** 对潜在空间的系统性探索，清晰地证明了我们的模型学习到了一个部分**解耦**的特征表示。它能够将抽象的随机向量映射到具体的、可解释的视觉特征上，例如**宏观的类别、微观的纹理、视角**等。
- **可控内容生成的潜力:** 这种解耦特性是实现可控内容生成的关键。通过在潜在空间中进行向量插值或算术运算，理论上可以实现对生成图像内容的精确编辑（如"穿靴子的凉鞋"），这在创意设计、数据增强等领域具有不可估量的应用价值。

未来的工作可以探索更先进的GAN变体（如WGAN, StyleGAN），以实现更稳定的训练和更高程度的特征解耦。