半监督时装图像识别

一、项目简介

本项目旨在探索半监督学习的实现方法，通过融合无监督的变分自编码器（VAE）与监督学习的分类器，构建了一个半监督学习的网络架构（Semi-SupVAE）。任务核心在于利用有限的标签数据和无标签数据共同训练网络，以提升图像分类的准确性和泛化能力。项目使用图像数据集部分带有类别为 10 的标签，部分则无标签。带标签数据为 5000 张 28x28 单通道数据集，我们通过数据增强策略，有效扩充了数据的多样性。在模型训练过程中，对有标签数据同时计算重建损失和分类损失，对无标签数据仅计算重建损失，以此实现半监督学习。最终，通过对比 CNN 模型和 Semi-SupVAE 模型在训练集、测试集以及评测集上的准确率和混淆矩阵，验证了 Semi-SupVAE 模型在全局性能上的优越性。

二、方法

1. 网络设计或算法描述

在探索半监督学习的实现方法中，我们从无监督的变分自编码器（VAE）模型获取灵感，并将分类器融入其中，成功构建了一个半监督学习的网络架构（Semi-SupVAE）。这一融合方式巧妙地结合了监督学习与无监督学习的优势，实现了半监督学习的目标。

在这个网络里，Encoder 模块承担着特征提取的关键任务，本质

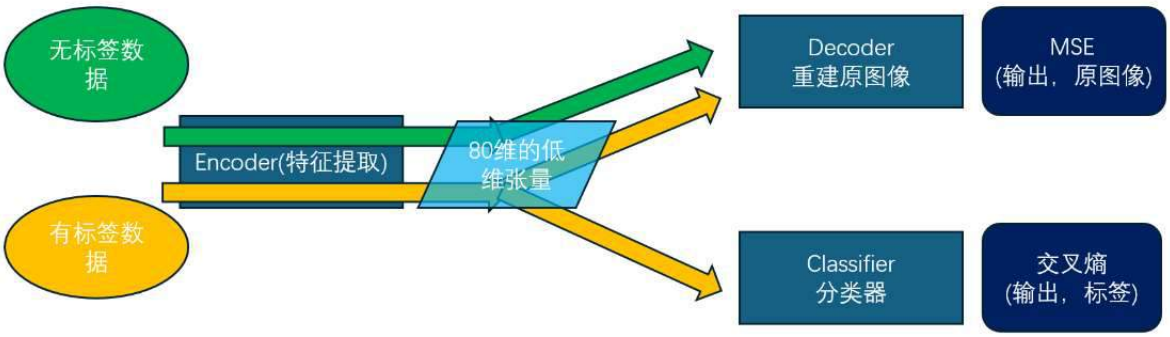
上这也是一种降维操作。特征提取旨在从繁杂的原始信息中提炼出关键要素，因此 Encoder 模块最终输出的是低维张量。对于本项目，我们设置的参数将原始图像的 784 维数据压缩至 80 维。值得一提的是，降维操作还为神经网络带来了额外的益处：显著提升其泛化能力，有效降低过拟合风险。这是因为低维度的数据无法提供充足的信息让神经网络去 “记忆”，从而促使网络学习到更具一般性的特征表示，而非特定数据的细节特征，进而避免了过拟合现象的发生。

我们深刻认识到，Encoder 模块所提取的特征必须具备良好的泛化能力，才能支撑整个网络的高效运行。基于此，我们将这些特征分别输入到以下两个神经网络中，开展多任务学习：

1. Decoder 模块（图像重建任务）：该模块利用 Encoder 输出的特征信息，致力于重建原始图像，这一过程与传统的 VAE 模型中的图像重建部分一样，并且此任务属于无监督学习范畴，无需依赖数据标签即可进行训练，它能够帮助网络学习到数据的内在结构和分布规律。
2. Classifier 模块（图像分类任务）：此模块利用相同的特征信息执行图像分类任务，这是我们在原有 VAE 模型基础上创新性添加的部分，该模块的训练依赖于有标签的数据，属于监督学习的范畴，通过利用标签信息，网络能够学习到不同类别图像的特征差异，从而实现准确的分类预测。

通过将 Decoder 模块和 Classifier 模块有机组合，我们成功构建了一个适用于半监督学习场景的神经网络。在实际训练过程中，对于无标签数据，我们仅利用其进行图像重建任务，即通过 Decoder 模块进行训练，其实根本目的还是训练 Encoder 网络；而对于有标签数据，我们则同时对图像重建和图像分类这两个任务进行训练，充分利用有限的标签信息，提升网络在分类任务上的准确性和泛化能力，从而实现半监督学习的高效性和优越性。

最终，网络的总体设计如下：

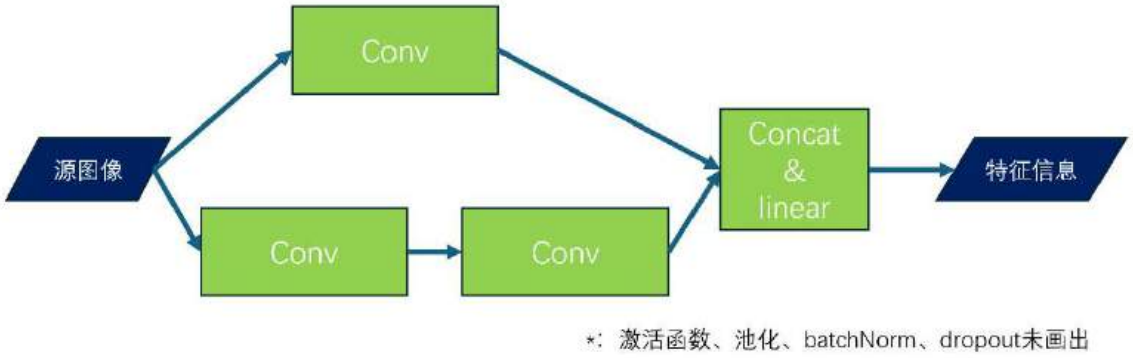


在 Encoder 网络的设计方面，我们借鉴了 YOLO 模型的理念，创新性地采用了将单层卷积与双层卷积的结果进行拼接的策略。首先对输入图像分别执行单层卷积和双层卷积操作，之后将这两个不同层次的卷积结果沿着特定维度进行拼接整合，最终通过一个线性层将拼接后的结果映射到低维空间，从而实现有效的特征提取与降维处理。

考虑到神经网络的输入为图像数据，其像素值的范围特点以及为了增强网络的非线性表达能力，我们选用了 ReLU 函数作为激活函数。ReLU 函数在处理图像数据时表现出良好的特性，它能够在保留图像

关键特征信息的同时，有效加速网络的训练过程，避免梯度消失问题，使得 Encoder 网络能够快速且准确地从输入图像中提取出具有代表性的低维特征向量，为后续的任务模块提供高质量的输入数据，进而提升整个网络模型的性能表现。

Encoder 网络设计图如下：

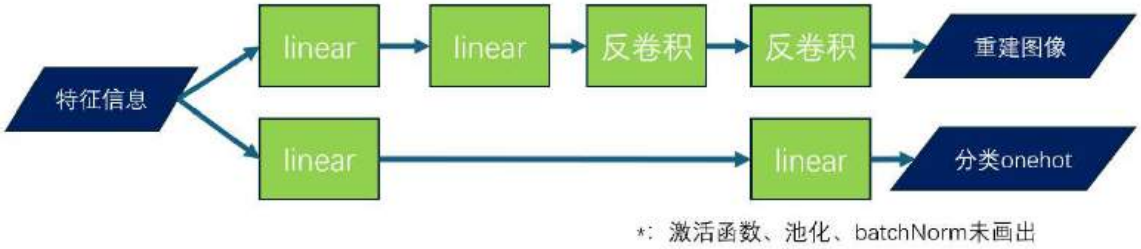


针对分类器部分，鉴于其输入已是经过特征提取后的关键信息，无需复杂的处理流程，仅需进行相对简单的变换操作，就能获取最终的分类结果。基于此，我们将分类器设计为一种精简的结构，仅由两个线性层组成。这种简约的设计能够在保证分类准确性的同时，显著降低计算复杂度，提高模型的运行效率。

而对于图像重建任务而言，Decoder 的设计与 Encoder 呈现出一种镜像对称的关系。具体来说，Decoder 首先将接收到的特征信息输入到线性层中，通过线性变换对特征进行初步的调整和扩展，恢复部分被压缩的信息维度和特征细节。在此基础上，再运用反卷积操作逐步复原出与原始图像尺寸相同、内容相近的图像数据，以此实现对原

始图像的高质量重建。这种镜像式的设计理念，使得 Decoder 能够充分利用 Encoder 提取的特征，为整个模型的半监督学习能力提供有力支撑，增强模型对无标签数据的利用效率和学习效果，进一步提升模型的泛化能力。

最终，多任务学习模块设计如下：



除了以上模型在架构方面的设计，为了显著增强模型的鲁棒性，使其能够更加稳定且精准地应对复杂多变的数据集以及各种潜在的噪声干扰，我们在卷积这一关键结构上引入了两种重要的正则化技术：dropout 和 batchNorm。

dropout 技术通过在训练过程中随机地将一部分神经元的输出设置为零，有效地避免了神经元之间的过度依赖，从而减少了模型的过拟合风险。这种随机失活的机制使得模型在每次训练迭代中都能够学习到不同的特征组合，进而增强了模型对数据的泛化能力，使其能够在面对未知数据时保持较高的准确性和稳定性。

与此同时，batchNorm（批量归一化）技术则致力于对每一层的输入数据进行归一化处理，使得数据的分布更加稳定且符合标准正态分布的特征。通过对每个批次的数据进行均值和方差的归一化操作，

batchNorm 能够有效地加速模型的训练过程，降低梯度消失或梯度爆炸的风险，并且提高模型对不同输入数据的适应性。这不仅有助于模型在训练阶段更快地收敛到更优的解，还能够在测试阶段保持一致的性能表现，进一步提升了模型的鲁棒性和可靠性。

2. 网络运行设置或算法运行设置

在模型训练阶段，针对数据集，我们首先实施了数据增强策略，通过对图像进行平移、翻转等操作，有效扩充了数据的多样性，从而提升模型对不同数据分布的鲁棒性。为确保模型在整体数据集上具备良好的性能表现，我们采用了批次大小为 32 的数据加载器（dataloader），以此平衡训练效率与模型收敛效果。

对于有标签的数据样本，将其输入到构建的网络中，同步计算重建损失和分类损失，并通过反向传播机制将误差回传，利用优化器对模型参数进行更新优化，使得模型在分类和重建任务上逐步收敛到更优的状态。而对于无标签的数据，仅计算其重建损失，并同样执行反向传播与优化器的参数更新操作，以增强模型对数据整体特征的学习和把握能力，尽管无标签数据无法直接指导分类任务的学习，但能够从数据分布的角度辅助模型学习到更通用的特征表示，提升模型的泛化性能。

得益于我们所采用的 Encoder-Decoder 架构设计，中间产生的特征向量处于低维空间，极大地降低了模型过拟合的风险，使得模型在

训练过程中能够持续学习而不易陷入过拟合困境。因此，我们可以持续不断地循环执行上述训练流程，即训练的轮次（epoch）理论上可以无限制地进行下去（实际操作中由用户手动决定何时终止程序运行）。在每一轮完整的训练结束后，立即启动测试流程，对模型在测试集上的性能进行评估。若模型在当前轮次测试中获得的准确率高于以往记录的最佳准确率，则保存当前神经网络的全部参数。

在测试流程中，对于每一个待测试的输入图像，同样先对其执行数据增强操作（包括平移、翻转等），随后将增强后的图像分别输入到神经网络中进行预测。由于网络输出的 one-hot 向量中每个元素的值代表着相应类别的预测概率，因此，对多次增强后的图像预测结果所对应的 one-hot 向量进行叠加，从概率统计的角度来看，这相当于一种投票机制，能够综合多次预测的结果，提升最终预测的准确性和稳定性。最后，选取概率值最高的元素所对应的下标，即可确定该图像的预测标签值，完成对单张图像的分类预测任务。

为对“final_x.npy”文件进行预测，我们采用与测试集相同的投票策略。即对其中每个样本先做与测试集一致的增强处理，再分别输入神经网络（参数设置为上文提到的测试效果最好的）得到 one-hot 向量预测结果，将这些结果叠加形成投票，最终确定每个样本的预测类别，从而生成针对该文件的最终预测值。

三、实验结果分析：

1. 对比方法：

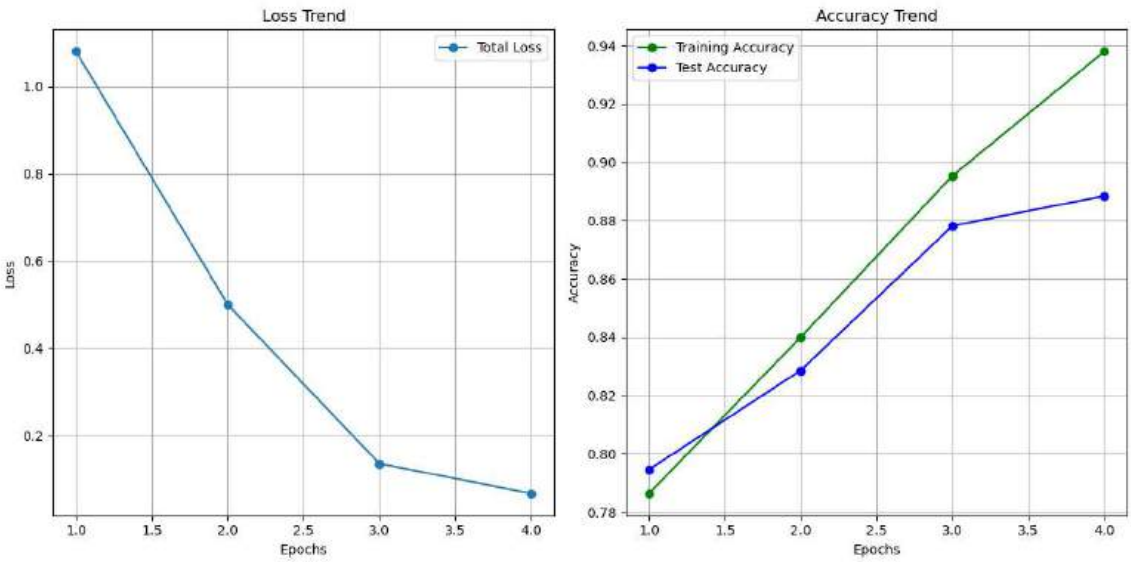
我们前后尝试了两种方法，一种是标准的 CNN 模型，另一种是我们创新的 Semi-supervised 模型，两种模型在同样的训练集训练后，同样的测试集测试下，CNN 模型准确率为 88.36，而 Semi-supervised 模型准确率达到 90.48，这证明了我们创新后的方法能更优越地解决时装图像识别的问题。

	Model	Training Accuracy	Testing Accuracy	Evaluation Accuracy (Highest)
0	CNN	93.79	88.83	88.36
1	Semi-supervised VAE	94.07	89.07	90.48

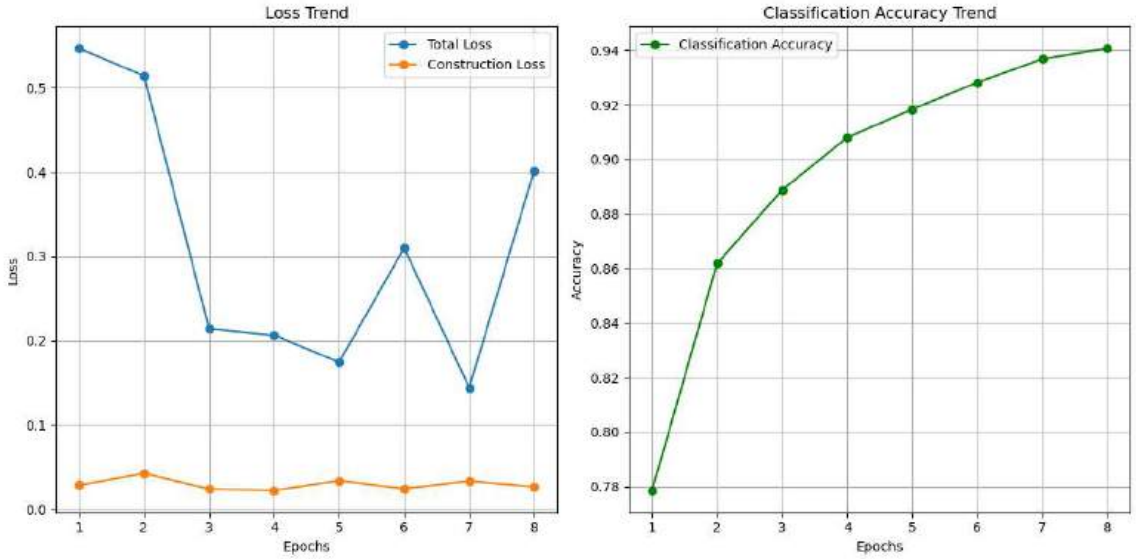
2. 评价指标一：

以准确率评估两种模型。先给出 CNN 和 Semi-SupVAE 模型各自在增强后的训练集上训练一个周期的各项指标。

CNN 模型：



Semi-SupVAE 模型：



在训练集、测试集以及最后的评测集上准确率对比：

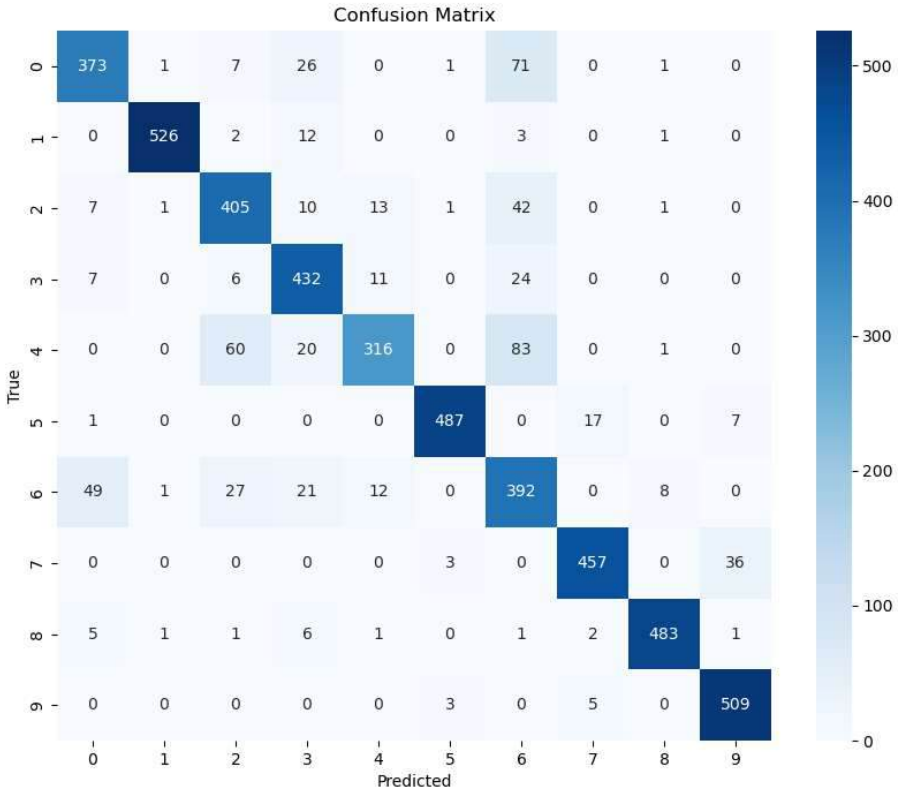
	Model	Training Accuracy	Testing Accuracy	Evaluation Accuracy (Highest)
0	CNN	93.79	88.83	88.36
1	Semi-supervised VAE	94.07	89.07	90.48

分析：可以发现，一个训练周期内，Semi-SupVAE 的收敛速度更快，无论在训练集还是测试集，效果均会更优。并且 Semi-SupVAE 在评测集上，最终能够实现更好的分类能力。

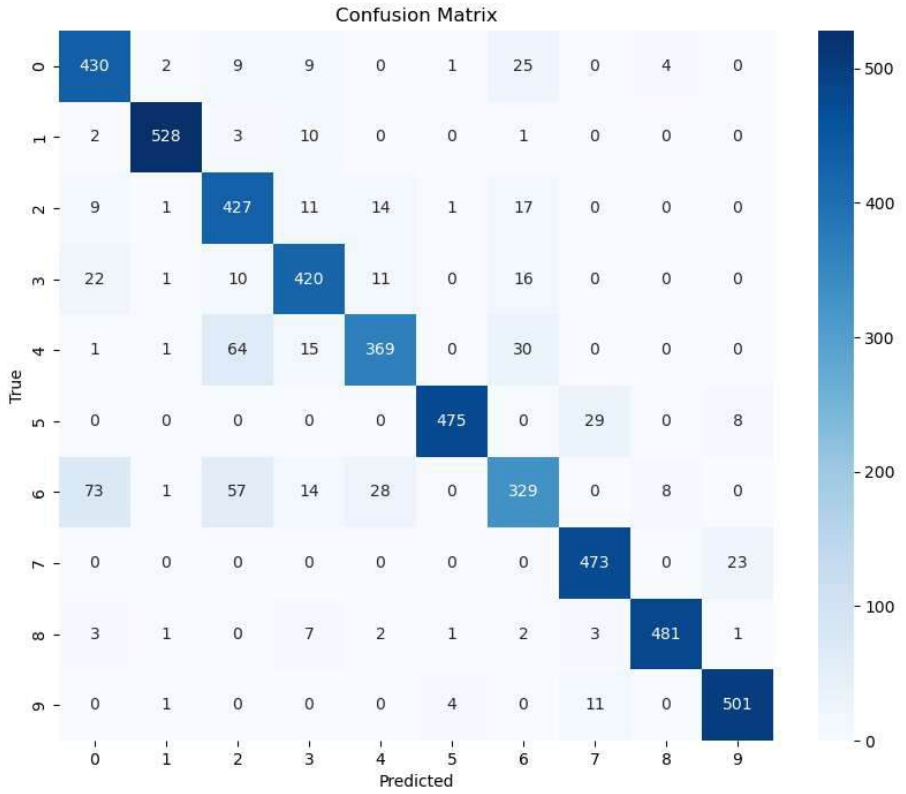
3. 评价指标二：

以混淆矩阵评估两种模型。先分别给出一个训练周期后，CNN 和 Semi-SupVAE 模型在测试集上的混淆矩阵。

CNN 模型：



Semi-SupVAE 模型：



分析：在一个训练周期下，Semi-SupVAE 模型在整体准确率上稍优，尤其在类别 0、1 和 9 上表现更好（正确分类数分别为 430、528、501）。然而，CNN 在类别 3 和类别 6 上错误分类更少，表现出对这些类别的更高鲁棒性（如类别 6 误分类率低于 Semi-SupVAE）。两模型均存在类别 4 和 6 间的显著混淆问题，但 CNN 对此的处理略优。综合来看，Semi-SupVAE 更适合追求全局性能，CNN 则更适合特定类别优化的场景。