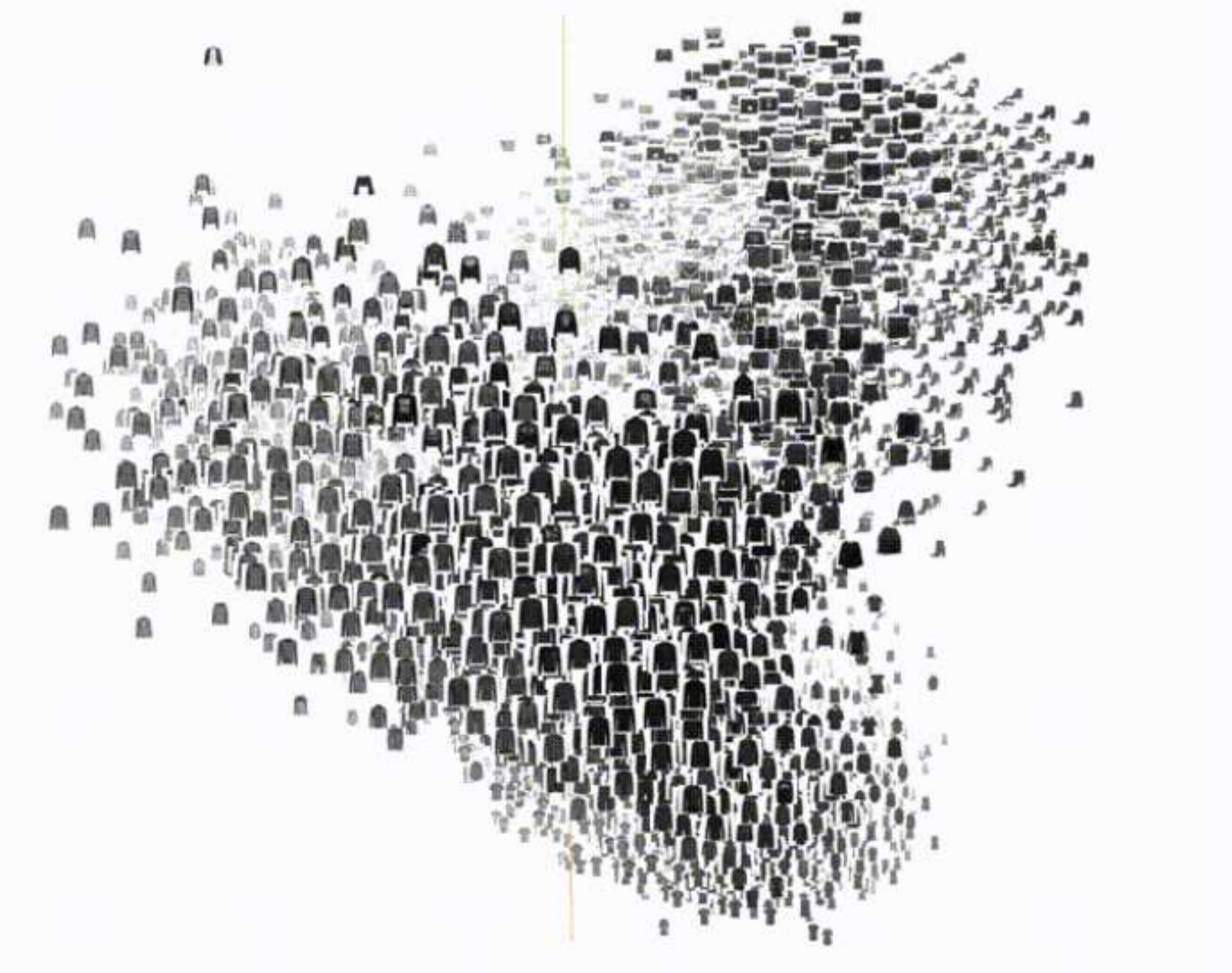
Final Project: Classification, Image Reconstruction and Model Analysis

Objective

- The objective of this project is to help you have a deep insight on machine learning, including an intuitive understanding of the feature and the model.
- Specifically, this project has two kinds of tasks, including the mandatory task and two optional tasks. You should complete **the mandatory task**, and **select one optional task.** All tasks should be implemented by yourself, instead of directly using any existing open source codes.
- As long as your code does not have a bug (e.g., the classification accuracy of the neural network trained by yourself is significantly lower than that of a neural network of similar complexity), we do not evaluate the project based on the accuracy, because this final project is not a competition of the accuracy.

1. Mandatory Task: Fashion-MNIST clothing classification

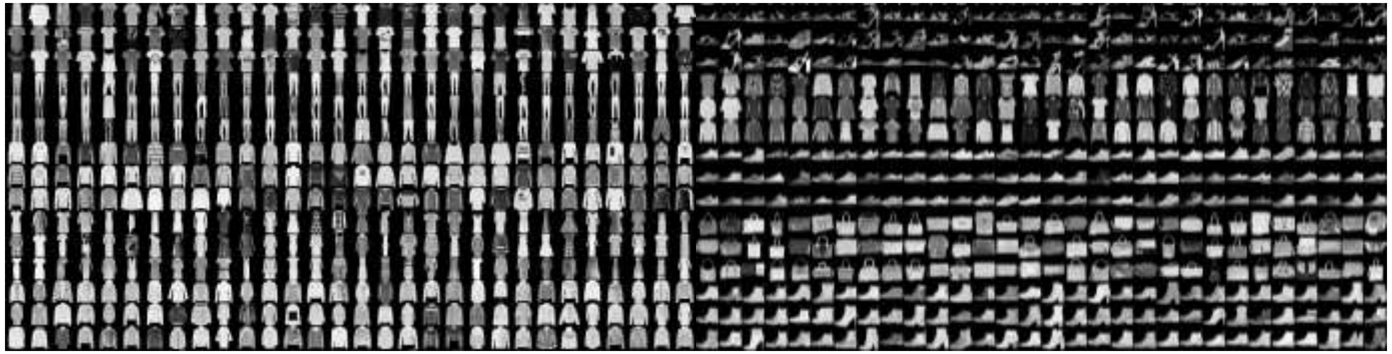


1.1 说明

- 此任务为必选任务，每位同学都需要完成本任务。
- 考虑到同学们拥有不同的计算资源，我们提供了两个选项来完成必选任务，你需要从以下两个选项中二选一执行。
  1. 如果你有充足的计算资源，请你阅读**标题1.2.1**下的内容并按照要求完成任务。
  2. 如果你缺少充足的计算资源，我们制定了数据集与网络结构的简化标准，让你能够使用CPU完成模型训练。请你阅读**标题1.2.2**下的内容并按照相应要求完成任务。

1.2.1 如果你拥有充足的计算资源

- 数据集：fashion-MNIST，请从canvas系统上下载压缩文件 data_required_opt1.zip，压缩包中包含数据集文件以及相应的数据接口，请在此基础上完成你的任务。
- 在fashion-MNIST 数据集当中，有60000个训练样本，10000个测试样本，每个样本是一张大小为32×32 的灰度图像，对应一个确定的类别（图像原大小为28×28，在提供的数据集中，已经进行插值放大到32×32），数据集中总共有十个类别。

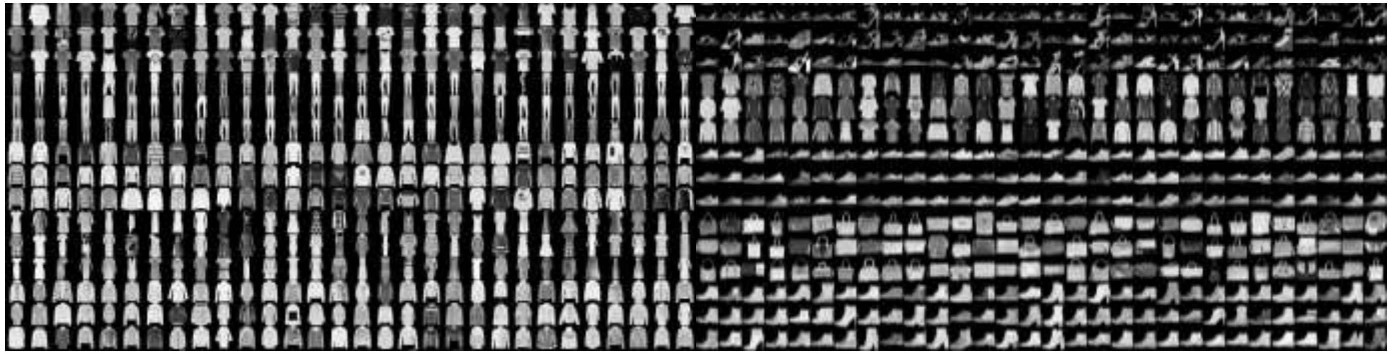


1.2.1 如果你拥有充足的计算资源

- You should design one neural network by yourself. Specifically, each designed neural network should contain 12-35 layers, including convolutional layers, ReLU layers, Batch Normalization layers, fully connected layers, and maxpooling layers.
  - You **should not directly** use classical neural networks (including but not limited to VGG-11/16/19, AlexNet, ResNet-18/24/32/36/44/56/102, DenseNet, GoogLeNet, and InceptionNet). Moreover, you should not design a new neural network by just adding or removing several layers from the above classical neural networks.
- You should train your designed neural networks on the dataset.
- You should use both **PCA** and **t-SNE** to visualize features on the designed neural network.

1.2.2 如果你缺少充足的计算资源

- 数据集：简化的fashion-MNIST，请从canvas系统上下载压缩文件 data_required_opt2.zip，压缩包中包含数据集文件以及相应的数据接口，请在此基础上完成你的任务。
- 在简化的fashion-MNIST 数据集当中，有1,000个训练样本，1,000个测试样本，每个样本是一张大小为32×32 的灰度图像，对应一个确定的类别（图像原大小为28×28，在提供的数据集中，已经进行插值放大到32×32），数据集中总共有十个类别。



1.2.2 如果你缺少充足的计算资源

- 你需要阅读技术博客《LeNet详解》，理解LeNet的结构设计，并实现 LeNet 网络。
- 你需要参考 LeNet 的网络结构，以此为基础设计一个3-7层的浅层神经网络，网络包含卷积层，池化层，激活层等。
- 你需要在数据集上训练两个神经网络，即 LeNet 以及你设计出来的网络，并分别给出训练以及测试结果。
- 你需要同时使用**PCA**以及**t-SNE**两个算法，可视化Lenet的中层输出特征，以及可视化你设计出来的网络的中层输出特征。

1.3 Feature Visualization

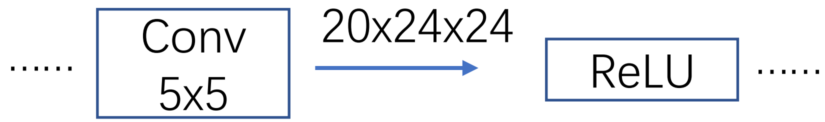
- **Principal component analysis (PCA):** PCA projects each data point onto only a few principal components to obtain lower-dimensional data, while preserving as much of the variance of data as possible (in Section 1 of lecture 6).
  - Specifically, given n samples X $\in \mathbb{R}^{n\times 784}$and a trained neural network g, let $g(X) \in \mathbb{R}^{n\times p}$denote the feature map in a specific intermediate layer of these input samples. You should use PCA to transform $g(X)$ to $g'(X) \in \mathbb{R}^{n\times 2}$ or $g'(X) \in \mathbb{R}^{n\times 3}$.
  - You should select at least three layers of the aforementioned network on page 4 to conduct PCA, and visualize $g'(X)$. The selected layers should at least include a convolutional layer, a fully-connected layer, and the final layer of the neural network.
  - The data matrix X should contain data samples from 10 classes, and each class should contain more than 15 samples, i.e. $n \geq 150$.

1.3 Feature Visualization

- **Stochastic Neighbor Embedding (t-SNE):** t-SNE uses lower dimensional vectors so as to preserve relationships that are in higher dimensional (in Section 2 of lecture 6).
  - Specifically, given n samples X $\in \mathbb{R}^{n\times 784}$ and a trained neural network g, let let $g(X) \in \mathbb{R}^{n\times p}$ denote the feature map in a specific intermediate layer of these input samples. You should use t-SNE to transform $g(X)$ to $g'(X) \in \mathbb{R}^{n\times 2}$ or $g'(X) \in \mathbb{R}^{n\times 3}$.
  - You should select at least three layers of the aforementioned network on page 4 to conduct t-SNE, and visualize $g'(X)$. The selected layers should at least include a convolutional layer, a fully-connected layer, and the final layer of the neural network.
  - The data matrix X should contain data samples from 10 classes, and each class should contain more than 15 samples, i.e. $n \geq 150$.

1.4 Requirements of the report for the mandatory task

- You should introduce architectures of neural networks designed by yourself and explain why you use such an architecture. It would be better to draw a diagram, which is shown as follows. Specifically, you should clarify the size of the feature map of each layer and hyper-parameters used to train the designed neural network.

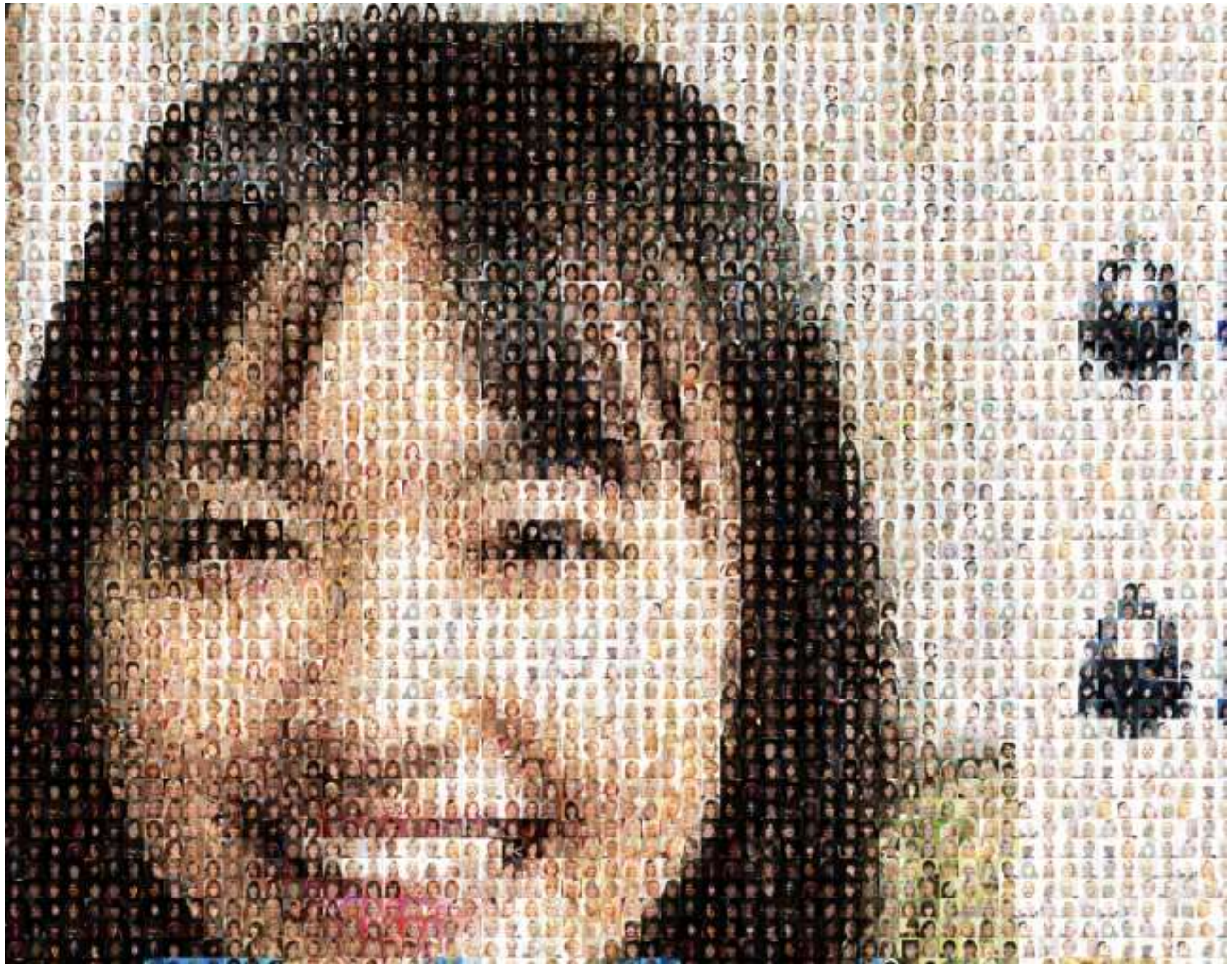


- You should report two curves for the training loss and the testing loss during the training process, respectively.

- You should report two curves for the training accuracy and the testing accuracy during the training process, respectively.

- You should use both PCA and T-SNE to visualize features. You should observe these visualization results and try to summarize some empirical conclusions. The summary of some empirical conclusions is an open problem.

2. Optional task 1: Image reconstruction



2.1 说明

- 本任务（图像重建）为选做任务一，每位同学需要在两个选做任务（图像重建以及解释深度神经网络）中选择一个任务完成。
- 考虑到同学们拥有不同的计算资源，我们提供了两个选项来完成选做任务一，你需要从以下两个选项中二选一执行。
    1. 如果你有充足的计算资源，请你阅读**标题2.2.1**下的内容并按照要求完成任务。
    2. 如果你缺少充足的计算资源，我们制定了数据集的简化标准，让你能够使用CPU完成模型训练。请你阅读**标题2.2.2**下的内容并按照相应要求完成任务。

2.2.1 如果你拥有充足的计算资源

- 数据集： Labeled Faces in the Wild (LFW)，请从canvas系统上下载压缩文件data_optional1_opt1.zip，压缩包中包含数据集文件以及相应的数据接口，请在此基础上完成你的任务。
- LFW 数据集包含了13000张以上从互联网上收集到的人脸图像样本。每一个样本是一张250× 250的RGB图像。在LFW数据集中，有1680 人对应两个或两个以上图像样本。

2.2.1 如果你拥有充足的计算资源

- You should design and train a VAE on the dataset.
- You should use Linear interpolation to generate images with specific properties.
    - For example, given two output features of the encoder, i.e., z_1 and z_2, the decoder takes z_1 as the input to generate a face image of a woman, and takes z_2 as the input to generate a face image of a man. You can use Linear interpolation to obtain a new feature $z = \alpha z_1 + (1 - \alpha) z_2, \alpha \in (0,1)$. Then, the decoder takes z as the input to generate a new face image. You are required to do experiments with different values of α, e.g., α=0.2, 0.4, 0.6, 0.8, so as to obtain a set of face images.

Image generated from z_1

Image generated from z

Image generated from z_2

- You can conduct experiments on any two categories (*e.g., male and female,* old and young)

2.2.2 如果你缺少充足的计算资源

- 数据集： 简化的Labeled Faces in the Wild (LFW)， 请从canvas系统上下载压缩文件 data_optional1_opt2.zip， 压缩包中包含数据集文件以及相应的数据接口，请在此基础上完成你的任务。
- 简化的LFW 数据集包含了1000个从互联网上收集到的人脸图像样本。每一个样本是一张32× 32的RGB图像。（图像原大小为250×250，已经进行插值缩小为32×32）

2.2.2 如果你缺少充足的计算资源

- You should design and train a VAE on the Simplified dataset.
- You should use Linear interpolation to generate images with specific properties.
  - For example, given two output features of the encoder, i.e., z_1 and z_2, the decoder takes z_1 as the input to generate a face image of a woman, and takes z_2 as the input to generate a face image of a man. You can use Linear interpolation to obtain a new feature $z = \alpha z_1 + (1 - \alpha) z_2, \alpha \in (0,1)$. Then, the decoder takes z as the input to generate a new face image. You are required to do experiments with different values of α, e.g., α=0.2, 0.4, 0.6, 0.8, so as to obtain a set of face images.

Image generated from z_1

Image generated from z

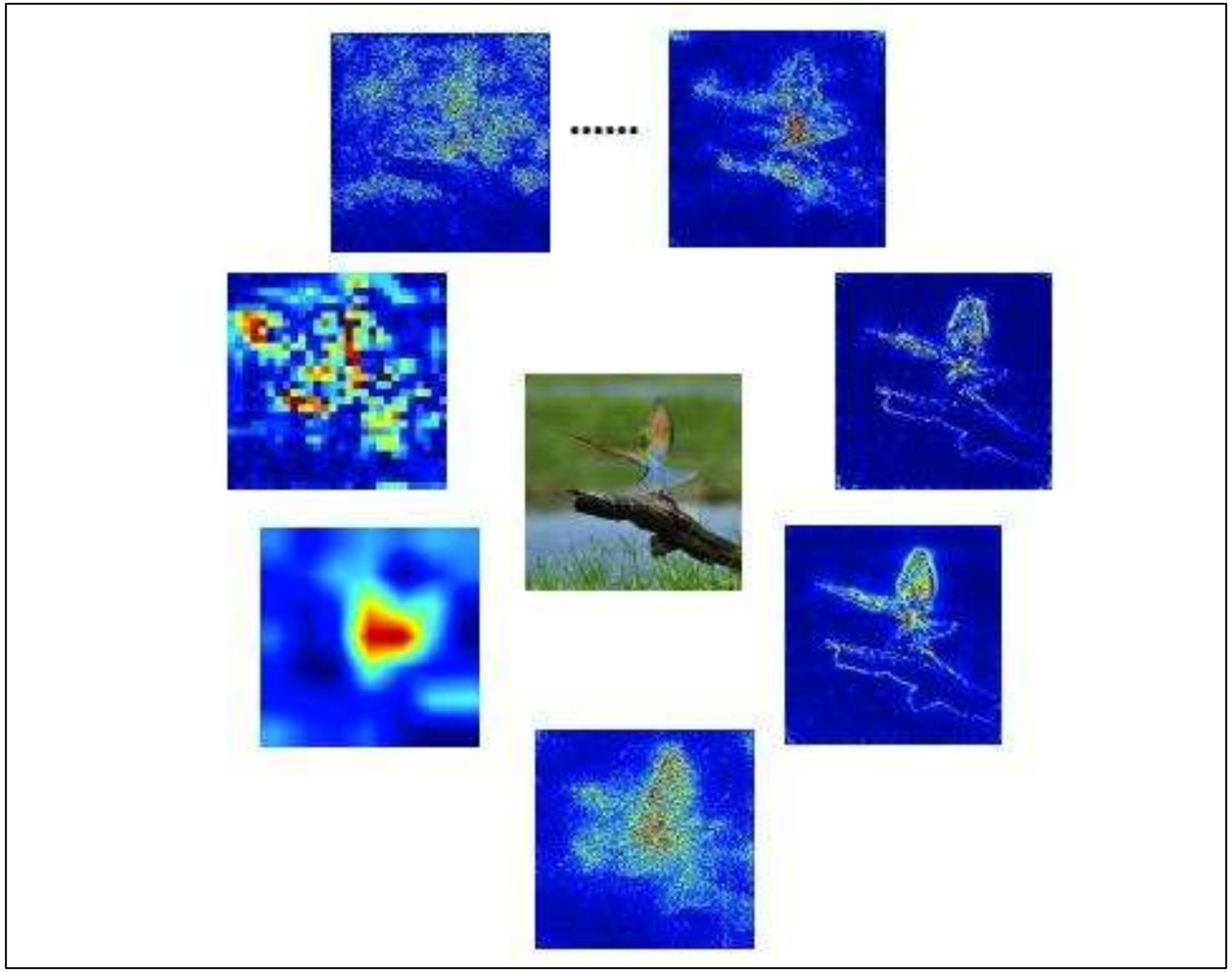
Image generated from z_2

  - You can conduct experiments on any two categories (*e.g., male and female,* old and young)

2.3 Requirements of the Report for Optional Task 1

- You should introduce the overall architecture of the VAE, which is designed by yourself. It would be better if you draw a diagram.
- You should write how you implement the VAE that is designed by yourself.
- You should visualize reconstructed images.
- You should use interpolation to generate images.
- **Note that we will not strictly compare the quality of generated images under similar network complexity, as long as there are not strange results that indicate a bug.**

3. Optional task 2: Visualization methods to explain DNNs



3.1 说明

本任务（解释深度神经网络）为选做任务二，每位同学需要在两个选做任务（图像重建以及解释深度神经网络）中选择一个任务完成。

3.2 Tasks for Optional Task 2

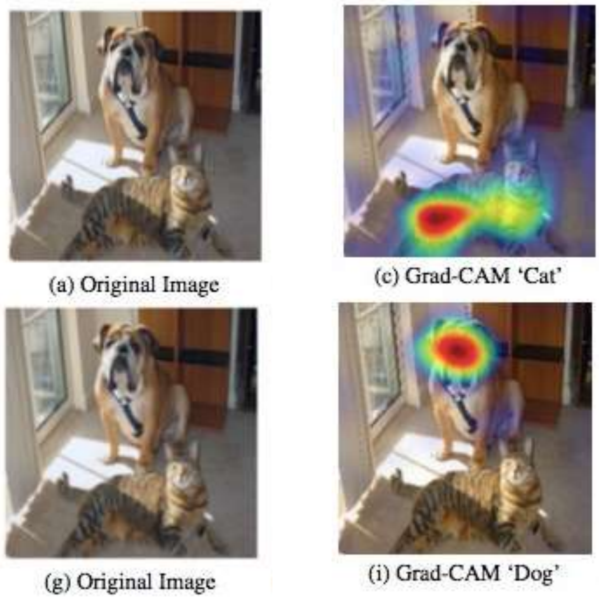
- Use visualization methods to explain at least two DNN models
  - You can download pre-trained DNN architectures directly.
    - ResNet-34 : https://download.pytorch.org/models/resnet34-b627a593.pth
  - Choose **at least two** visualization methods to explain models.
    - Optional task 2 requires you to read papers. These papers are easy to read.
    - Select **at least two** of Grad-CAM / Shapley value / Integrated Gradients to explain the pre-trained models.
    - Compare the visualization results of the visualization methods you chose for the pre-trained models, and analyze your visualizations.
    - Write the code by yourself, instead of using any open source code directly.

3.3 Requirements of the Report for Optional Task 2

- You should visualize the results of the visualization methods you chose for the pre-trained models.
- You should compare the visualization results and give an analysis of your visualizations.
- You should report the mathematical formula for the selected methods. Please explain the formula and the notations clearly.

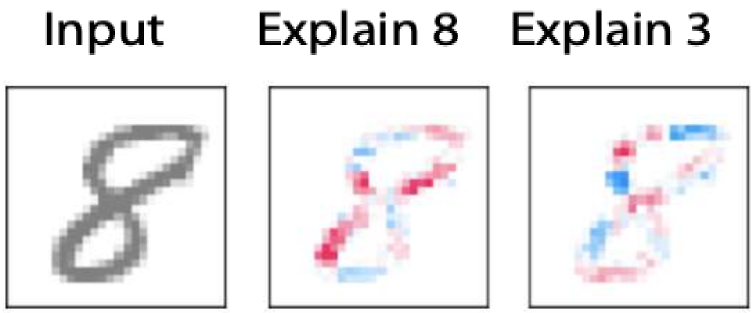
3.4 Grad-CAM

- **Grad-CAM**: Visualize the importance of regions in the input image.
- Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra, Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization, in arXiv:1610.02391



3.5 Shapley value

- **Shapley value** : Visualize the importance of individual pixels or regions in the input image.
- (Shapley value) Lloyd S Shapley. A value for n-person games. In: Contributions to the Theory of Games 2.28 (1953), pp. 307–317. 28. Princeton, NJ: Princeton University Press; 1957. p. 307--17.
- (Shapley value to visually explain the model) Please read **Section 3.1** of Hao Zhang, Jiayi Chen, Haotian Xue, Quanshi Zhang, Towards a unified evaluation of explanation methods without ground truth, in arXiv: 1911.09017.（This paper does not use the Shapley value to visually explain the model. Instead, this paper briefly introduces the Shapley value in Section 3.1. ）



3.6 Integrated Gradients

- Integrated Gradients: Visualize the importance of individual pixels in the input image.
- Mukund Sundararajan, Ankur Taly, Qiqi Yan, Axiomatic Attribution for Deep Networks, in arXiv:1703.01365

Original image

Integrated gradients

Submit

- Code & report
  - **Do not copy others’ codes or others’ reports**
    - **The teaching assistant will check the code and the report.**
  - **Make sure your code can run**
- Deadline
  - Wednesday in the 17th week (June 7)
- Please contact the teaching assistants if you have any questions.