# 基于神经网络的图像分类

## 1 实验简介

自深度学习兴起以来，神经网络在各个应用领域都取得了令人瞩目的成就，特别是在图像分类领域，以 AlexNet [3], VGG [5], GoogLeNet [6], ResNet [2], ViT [1] 等为代表的深度神经网络方法逐步取代了传统方法，相应的分类准确度逐步接近和超过人类。

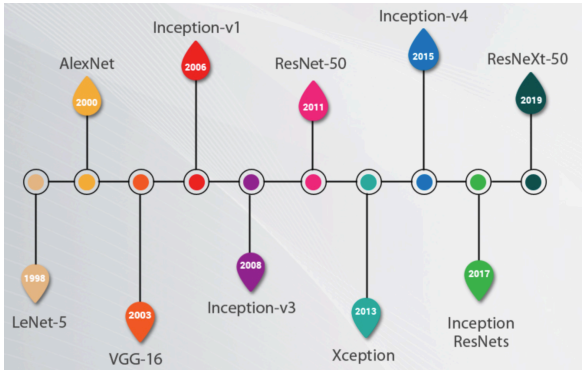


图 1: 深度神经网络发展示意图

### 1.1 图像分类

图像分类旨在学习一个分类器能对输入图像分类到已有类别（记类别数为 $k$），通常是输出一组预测概率，如下所示：

$$f_\theta(x) = \mathbf{p} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_k \end{pmatrix} = \begin{pmatrix} \frac{\exp(z_1)}{\sum_{i=1}^{k}\exp(z_i)} \\ \frac{\exp(z_2)}{\sum_{i=1}^{k}\exp(z_i)} \\ \vdots \\ \frac{\exp(z_k)}{\sum_{i=1}^{k}\exp(z_i)} \end{pmatrix} \tag{1}$$

其中 $p_i$ 表示分类器预测为第 $i$ 类的概率（一般用 softmax 来获取），$\theta$ 表示分类器的参数。对于基于深度学习的图像分类来说，我们通常需要训练集，它包含了很多的图像-标签

对 $\{(x_i, y_i)\}_{i=1}^N$，从而以最小化损失的方式，通过梯度反向传播更新网络参数。常用的损失函数包括 Cross Entropy，MSE，MAE 等，具体形式如下：

$$CE = -\frac{1}{N}\sum_{i=1}^{N}\log p_{i,y_i},\ MSE = \frac{1}{N}\sum_{i=1}^{N}\|\mathbf{p}_i - \mathbf{e}_{y_i}\|_2^2,\ MAE = \frac{1}{N}\sum_{i=1}^{N}\|\mathbf{p}_i - \mathbf{e}_{y_i}\|_1, \tag{2}$$

其中 $\mathbf{p}_i$ 表示第 $i$ 张图像的预测概率，$p_{i,j}$ 表示第 $i$ 张图像预测为第 $j$ 类的概率，$\mathbf{e}_i$ 表示第 $i$ 个坐标值为 1 的单位向量，$\|\mathbf{z}\|_2 = \sqrt{\sum_{i=1}^k z_i^2}$ 为 2 范数，$\|\mathbf{z}\|_1^1 = \sum_{i=1}^k |z_i|$ 为 1 范数。

常用的评价指标包括 top1 acc 和 top5 acc，本实验只考虑 top1 acc，如下所示：

$$acc = \frac{\sum_{i=1}^{N}\mathbb{I}(\arg\max_j p_{i,j} = y_i)}{N}. \tag{3}$$

# 2 实验目的

1. 熟悉并掌握图像分类的基本原理和代码实现；
2. 熟悉并掌握各时期分类网络运算原理和相关应用，加深对神经网络的了解和认识；
3. 了解手写数字识别数据集 MNIST [4]，能够完整实现手写数字识别及相关扩展；

# 3 实验内容

## 3.1 文献阅读 (30%)

1. 阅读相关文献，了解图像分类方面的神经网络；
2. 整理分类网络的发展历程，撰写报告，并绘制相关时间线，如图1所示；

## 3.2 手写计算器 (50%)

1. 了解 MNIST 数据集，并对数据集进行扩充，添加 +，−，×，÷，(，) 这些运算符，对应数据集可以点击获取 (密码 jods)；
2. 设计简单的卷积网络，实现一个手写体识别器，包括数字和运算符的识别；
3. 报告完整实验流程，训练参数，实验细节等；
4. 展示手写数字以及运算符号的分类结果；
5. 描绘训练过程和测试过程，包括但不限于训练损失曲线和测试准确度曲线；
6. 对比不同损失的实验效果，包括但不限于 CE，MSE，MAE；
7. 补充：实现完整的手写体识别四则运算，可以是网页服务或手机 APP。

### 3.3 对抗鲁棒性 (20%)

1. 熟悉并掌握对抗鲁棒性的原理和方法，参考https://adversarial-ml-tutorial.org/introduction/；
2. 设计实验获取手写体识别器的对抗样本；
3. 报告完整实验流程，训练参数，实验细节等；
4. 展示原样本，对抗样本和所添加的扰动，以及攻击前后的预测概率。

## 4 实验要求

1. 不限制语言和深度学习框架，推荐 python+tensorflow/pytorch；
2. 提交代码和实验报告，代码要完整有注释，报告要清晰且简洁；
3. 鼓励大家尝试其他数据集和其他方法，如有创新，可适当加分；
4. 如计算资源受限，可自行划分数据集，并在报告中作详细说明，实验室也可适当提供 GPU 资源。

注：严禁抄袭，如有雷同，成绩记 0!!!

## 参考文献


[1] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[3] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.

[4] Yann LeCun, Corinna Cortes, and CJ Burges. Mnist handwritten digit database. *ATT Labs [Online]. Available: http://yann.lecun.com/exdb/mnist*, 2, 2010.

[5] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[6] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.