

# AI 生成的图像能帮助图像识别吗？

GISLab 暑期短训班 2025



陈振源  
浙江大学地球科学学院

2025

[bili\\_sakura@zju.edu.cn](mailto:bili_sakura@zju.edu.cn)

# 目录

- ▶ 什么是图像分类?
- ▶ 我们如何改进图像分类? (数据增强)
- ▶ 真实案例: 灾害图像
- ▶ 项目: AI 生成的图像能帮忙吗?

# 什么是图像分类？

- ▶ 计算机学习识别图片中的内容（如猫、狗、飞机等）。
- ▶ 我们用大量带标签的图片来教计算机。
- ▶ 目标：为新图片预测正确的标签。

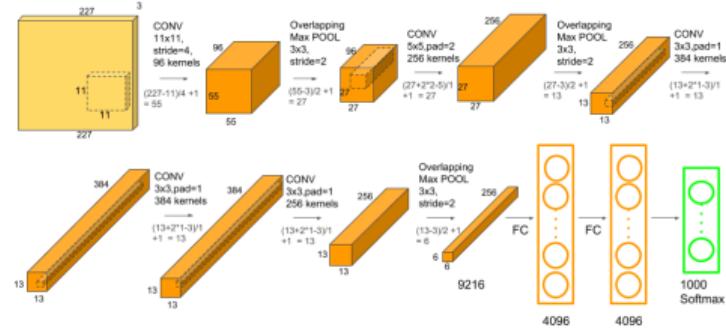


# 它是如何工作的？

神经网络是一种特殊的计算机程序，可以从大量示例图片中学习模式。随着时间推移，它们在区分图片方面会变得越来越好。



图：ILSVRC-2010 (Berg, Deng, and Fei-Fei, 2010) 测试图片及模型认为最可能的五个标签 (Krizhevsky, Sutskever, and Hinton, 2012)。



图：图像分类神经网络示例：AlexNet (Krizhevsky, Sutskever, and Hinton, 2012)。

# 现代模型：CLIP

- ▶ CLIP 学习将图片和文本进行匹配。
- ▶ 它可以通过理解描述来识别新事物。
- ▶ 不仅仅用于分类，还适用于多种任务。

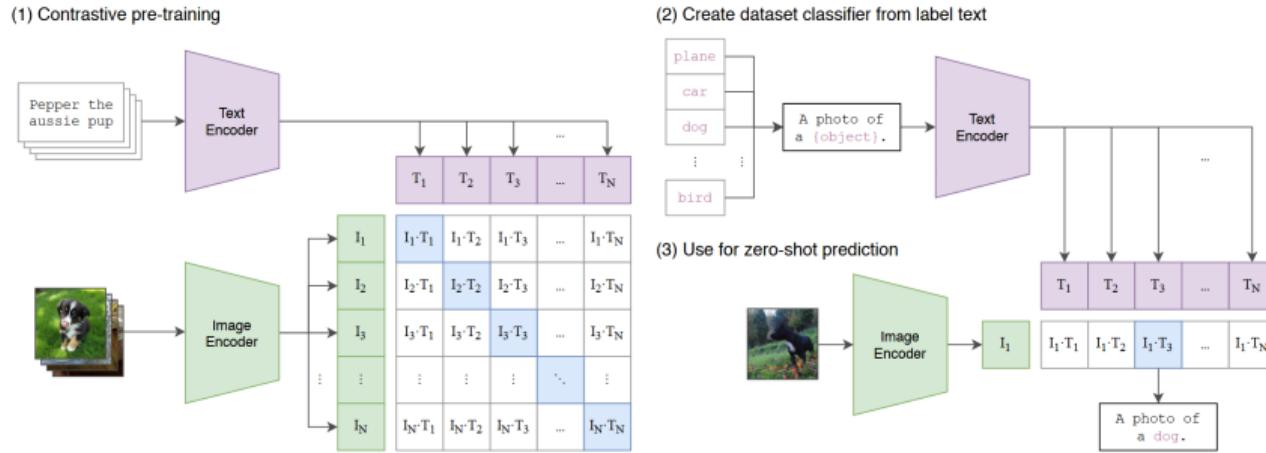
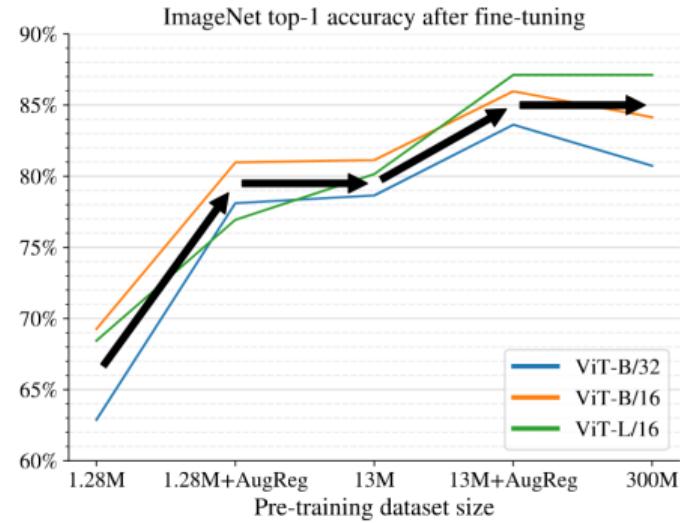


图: OpenAI CLIP ViT 总结 (Radford et al., 2021)。

# **我们如何改进图像分类? (数据增强)**

# 为什么要用数据增强？

- ▶ 有时我们没有足够的图片来训练一个好的模型。
- ▶ 数据增强意味着用已有图片生成新的图片。
- ▶ 它能帮助模型学得更好，减少错误。



图：适当的正则化和图像增强可以带来与数据集扩大十倍相似的提升效果。 (Steiner et al., 2022)

# 我们如何进行数据增强？

## 经典方法：

- ▶ 翻转、旋转、裁剪、改变颜色等。

## 现代方法：

- ▶ 将两张图片混合（Mixup）(Zhang et al., 2018)。
- ▶ 剪切并粘贴图片的一部分（CutMix）(Yun et al., 2019)。



软标签示例（CutMix）：

$$\text{cutmix\_label} = \lambda \cdot \text{label}_A + (1 - \lambda) \cdot \text{label}_B$$

示例：  $\lambda = 0.5$ ,  $\text{label}_A = [1, 0]$ ,  $\text{label}_B = [0, 1]$

$$\text{cutmix\_label} = 0.5 \times [1, 0] + 0.5 \times [0, 1] = [0.5, 0.5]$$

图：现代数据增强方法示意图。从左到右依次为：Mixup (Zhang et al., 2018), Cutout (DeVries and Taylor, 2017), CutMix (Yun et al., 2019)。

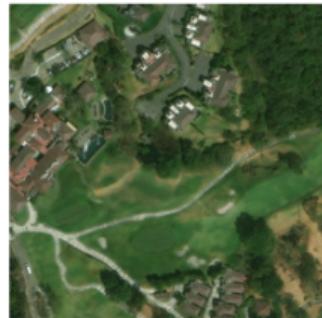
Zhang, et al. Mixup: Beyond Empirical risk minimization, ICLR, 2018.

Yun, et al. CutMix: Regularization Strategy to Train Strong Classifiers With Localizable Features, ICCV, 2019.

DeVries, et al. Improved Regularization of Convolutional Neural Networks with Cutout, arXiv, 2017.

# 用生成式 AI 制作新图片

- ▶ 生成式模型可以创造新的、逼真的图片。
- ▶ 我们可以用它们来生成更多训练数据。
- ▶ 例如：给一张“前”图片和描述，生成一张新的“后”图片。



+ suffer from  
volcano eruption

$C_T$

$$X_{post} \sim p(X|X_{pre}, C_T)$$

Generative Models



## **真实案例：灾害图像 (xBD)**

# xBD：大规模灾害损失数据集



xBD (Gupta et al., 2019) 是一个双时相遥感数据集，涵盖 19 种不同的灾害事件。上方为灾前影像，下方为灾后影像。从左到右依次为：哈维飓风、乔普林龙卷风、Lower Puna 火山喷发、巽他海峡海啸。

表：xBD 中的 19 个灾害事件。

灾害类型	灾害事件	事件日期
地震	墨西哥城地震	2017 年 9 月 19 日
森林火灾	葡萄牙森林火灾	2017 年 6 月 17–24 日
森林火灾	圣罗莎森林火灾	2017 年 10 月 8–31 日
森林火灾	卡尔森林火灾	2018 年 7 月 23–8 月 30 日
森林火灾	伍尔西火灾	2018 年 11 月 9–28 日
森林火灾	皮纳火灾	2018 年 11 月 25–12 月 2 日
火山	Lower Puna 火山喷发	2018 年 5 月 23–8 月 14 日
火山	危地马拉富埃戈火山喷发	2018 年 6 月 3 日
风暴	阿拉巴马州塔斯卡卢萨龙卷风	2011 年 4 月 27 日
风暴	密苏里州乔普林龙卷风	2011 年 5 月 22 日
风暴	俄克拉荷马州穆尔龙卷风	2013 年 5 月 20 日
风暴	马修飓风	2016 年 9 月 28–10 月 10 日
风暴	佛罗伦萨飓风	2018 年 9 月 10–19 日
洪水	尼泊尔、印度、孟加拉国季风	2017 年 8 月
洪水	哈维飓风	2017 年 8 月 17–9 月 7 日
洪水	迈克尔飓风	2018 年 10 月 7–16 日
洪水	美国中西部洪水	2019 年 1 月 3–5 月 31 日
海啸	印度尼西亚海啸	2018 年 9 月 18 日
海啸	巽他海峡海啸	2018 年 12 月 22 日

# 项目：AI 生成的图像能帮忙吗？

# 项目：你会做什么？

- ▶ 测试加入 AI 生成图片是否能帮助模型学习。
- ▶ 尝试三种方式：
  - ▶ 只用真实图片
  - ▶ 只用生成图片
  - ▶ 真实图片和生成图片混合
- ▶ 看哪种方式效果最好！

# 项目：数据集细节与如何生成图片

- ▶ 每种灾害类型使用 100 张真实图片（共 6 类，600 张图片）。
- ▶ 每类用 AI 生成 100 – 400 张新图片。
- ▶ 尝试不同配比：1:1、1:2、1:3、1:4（真实:生成）。
- ▶ 使用商用 AI 模型（如 GPT-Image-1 (OpenAI, 2025)、Gemini 2.5 Pro (Gemini Team, Google, 2025)、SeedEdit 3.0 (Wang et al., 2025)）生成图片。
- ▶ 输入：一张“前”图片和简短描述（如“让它遭受洪水”）。
- ▶ 输出：一张新的、逼真的“后”图片。

# 用哪些模型与如何衡量效果

## ▶ 尝试这些模型：

- ▶ OpenAI CLIP (Radford et al., 2021)
- ▶ Google SigLip/SigLip2 (Tschanne et al., 2025; Zhai et al., 2023)
- ▶ 以上模型均采用 ViT 结构 (见 [CLIP 训练示例](#) 和 [ViT 教程](#))。

## ▶ 如何衡量效果：

- ▶ **准确率**: 正确答案的百分比
- ▶ **F1 分数**: 平衡精确率和召回率
- ▶ **混淆矩阵**: 展示哪些类别容易混淆
- ▶ 始终在模型未见过的真实图片上测试。
- ▶ 用简单的柱状图或折线图对比结果。

Radford, et al. Learning Transferable Visual Models From Natural Language Supervision, ICML, 2021.

Tschannen, et al. (2025). SigLIP 2: Multilingual Vision-Language Encoders with Improved Semantic Understanding, Localization, and Dense Features.

Zhai, et al. Sigmoid Loss for Language Image Pre-Training, 2023.