

I I. Introduction:

這次 LAB 的目的主要在於熟悉 Data Augmentation 以及 Data Sampler 相關 API 的使用，學會在面對 Data imbalanced 或是資料過少的問題時，如何有效利用這些方法去提升模型的準確率。

II. Experiment setup

同 Lab1

I III. Result

2-1-1:

以下是我結合 imgaug 以及 torch.transform 的 data Augmentation 方式

```
class ImgAugTransform:
    def __init__(self):
        self.aug = iaa.Sequential([
            iaa.Resize((224, 224)),
            iaa.Sometimes(0.25, iaa.GaussianBlur(sigma=(0, 3.0))),
            iaa.Fliplr(0.5),
            iaa.Affine(rotate=(-20, 20), mode='symmetric'),
            iaa.Sometimes(0.25,
                           iaa.OneOf([iaa.Dropout(p=(0, 0.1)),
                                       iaa.CoarseDropout(0.1, size_percent=0.5)])),
            iaa.AddToHueAndSaturation(value=(-10, 10), per_channel=True)
        ])

    def __call__(self, img):
        img = np.array(img)
        return self.aug.augment_image(img)

tfs = transforms.Compose([
    ImgAugTransform(),
    transforms.ToTensor(),
    transforms.Normalize((0.5, 0.5, 0.5), (0.5, 0.5, 0.5))
])
```

其中 16 筆資料轉換後的結果



2-1-2

我以 100 為基準(表示不變)，超過 100 根據相對於 100 的倍率增加

image_filenames 裡的資料(隨機採樣並複製)，低於 100 則刪除，使

用的權重是原本的例子

wt = [125, 80, 25, 100, 200, 800, 80, 60, 40, 150, 1000]

最後結果:

```
-----  
Dataset - food11re/skewed_training  
class_name      |bf. loading      |af. loading  
Bread            |994              |1242  
Dairy_product    |128              |102  
Dessert          |1500             |375  
Egg              |591              |591  
Fried_food       |508              |1016  
Meat             |1325             |10600  
Noodles          |132              |106  
Rice             |84               |50  
Seafood          |513              |205  
Soup            |1500             |2250  
Vegetable_fruit |35               |350  
-----
```

2-2

以每個 class 的 sample 數量的倒數作為那個 class 的權重，最後就能

得到平均分配的結果，我採樣 10000 筆資料，上半為

WeightedRandomSampler，下半為 RandomSampler，可以發現

WeightedRandomSampler 後的結果是平均分配的，而

RandomSampler 的比例還是原本每個 class 當中 sample 數量的比例

Dataset - food11re/skewed_training		
class_name	bf. loading	af. loading
Bread	994	949
Dairy_product	128	921
Dessert	1500	897
Egg	591	891
Fried_food	508	936
Meat	1325	900
Noodles	132	842
Rice	84	970
Seafood	513	890
Soup	1500	893
Vegetable_fruit	35	911

Dataset - food11re/skewed_training		
class_name	bf. loading	af. loading
Bread	994	1308
Dairy_product	128	173
Dessert	1500	2069
Egg	591	832
Fried_food	508	703
Meat	1325	1801
Noodles	132	169
Rice	84	130
Seafood	513	735
Soup	1500	2029
Vegetable_fruit	35	51

2-3

來不及 train 完= ~=(LAB1 用 resnet18 當 pretrained model 有到 86%)