



自定义数据集实战

主讲：龙良曲

Pokemon Go!



Pokemon Dataset



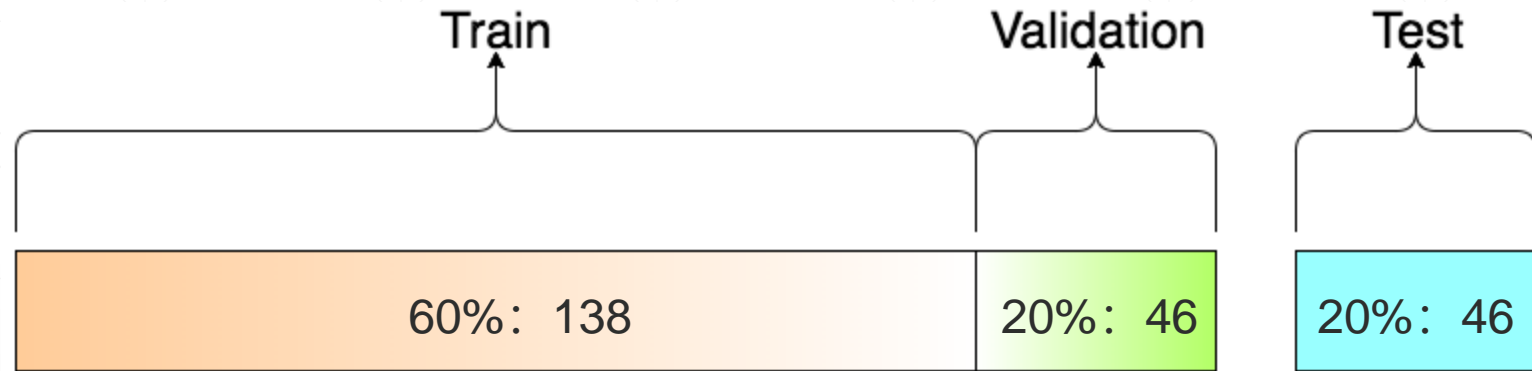
Download

- 链接: https://pan.baidu.com/s/1V_ZJ7ufjUUFZwD2NHSNMFw
- 提取码: dsxl



Splitting

- 皮卡丘: 234
- 超梦: 239
- 杰尼龟: 223
- 小火龙: 238
- 妙蛙种子: 234



4 steps

- Load data
 - Build model
 - Train and Test
 - Transfer Learning
-

Step1.Load data

- Inherit from `torch.utils.data.Dataset`
 - `__len__`
 - `__getitem__`
-

Custom Dataset



```
class NumbersDataset(Dataset):
    def __init__(self, tranining=True):
        if tranining:
            self.samples = list(range(1, 1001))
        else:
            self.samples = list(range(1001, 1501))

    def __len__(self):
        return len(self.samples)

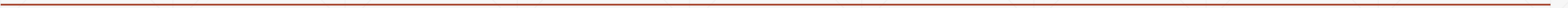
    def __getitem__(self, idx):
        return self.samples[idx]
```


Preprocessing

- Image Resize
 - 224x224 for ResNet18
 - Data Argumentation
 - Rotate
 - Crop
 - Normalize
 - Mean, std
 - ToTensor
-

Step2.build model

- Inherit from base class
- Define forward graph

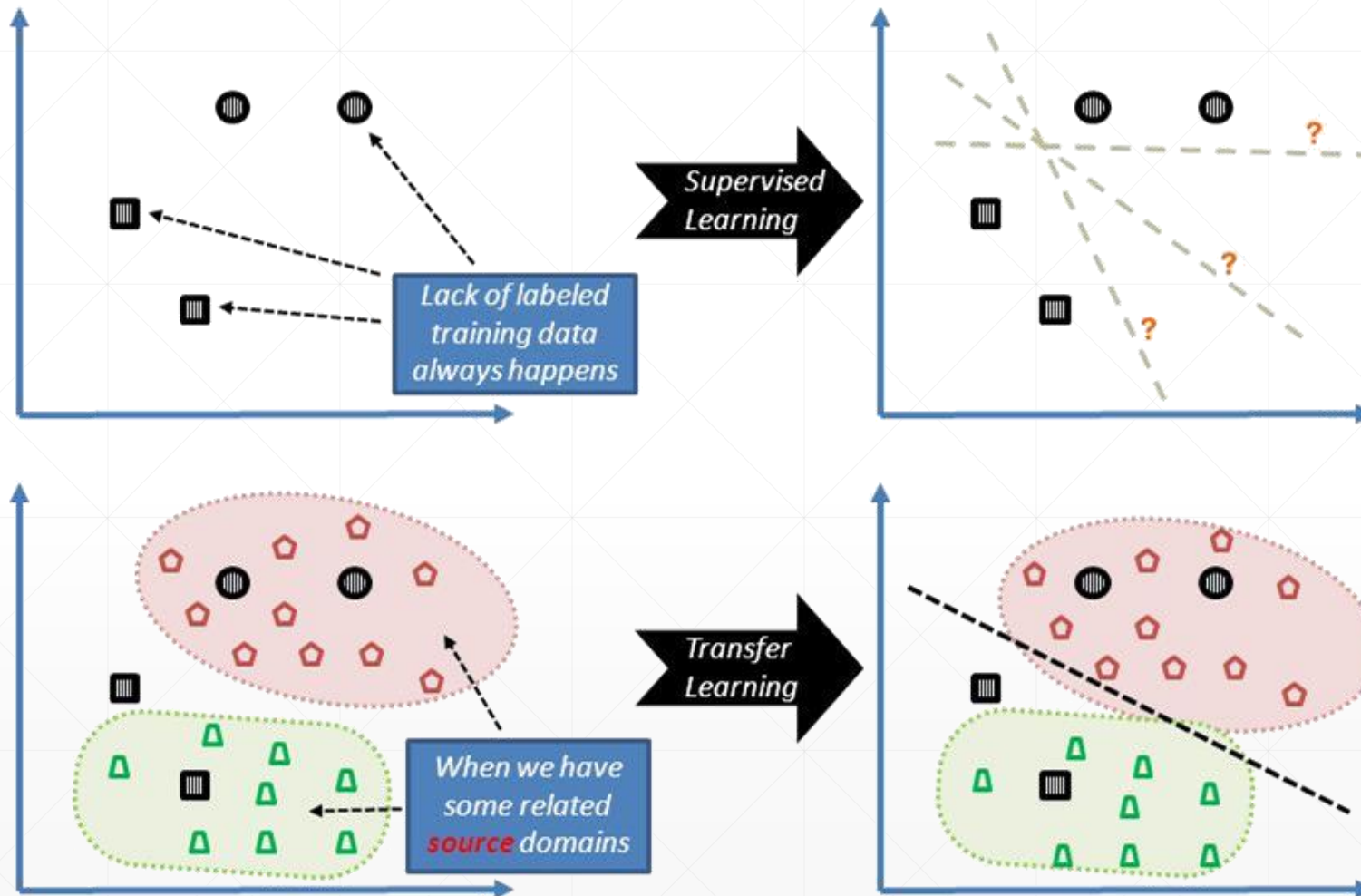


Step3.Train and Test

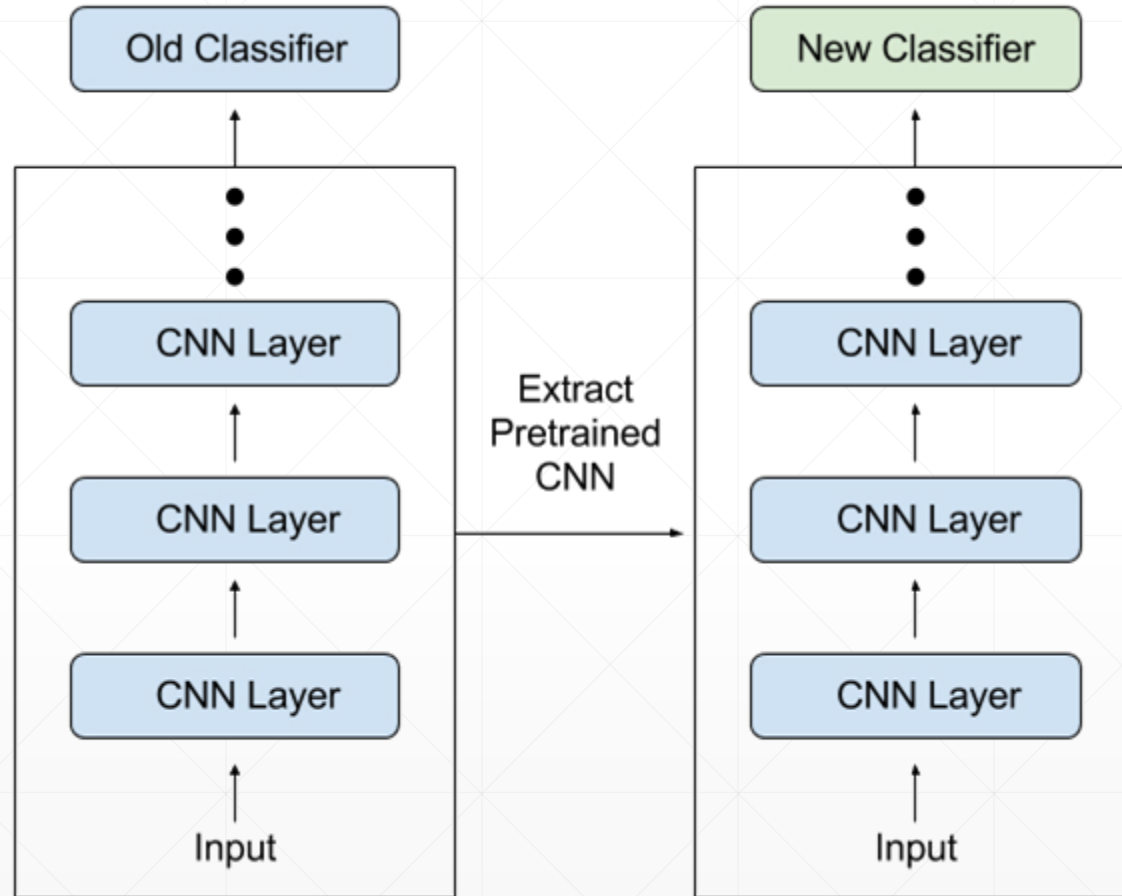


```
for epoch in range(epochs):  
  
    train(train_db)  
  
    if epoch%10==0:  
        val_acc = evaluate(val_db)  
  
        if val_acc is the best:  
            save_ckpt()  
        if out_of_patience():  
            break  
  
load_ckpt()  
test_acc = evaluate(test_db)
```

Step4.Transfer learning



In details



In Conclusion

- Load custom data
 - Train from scratch
 - Transfer learning
-

下一课时

Thank You.
