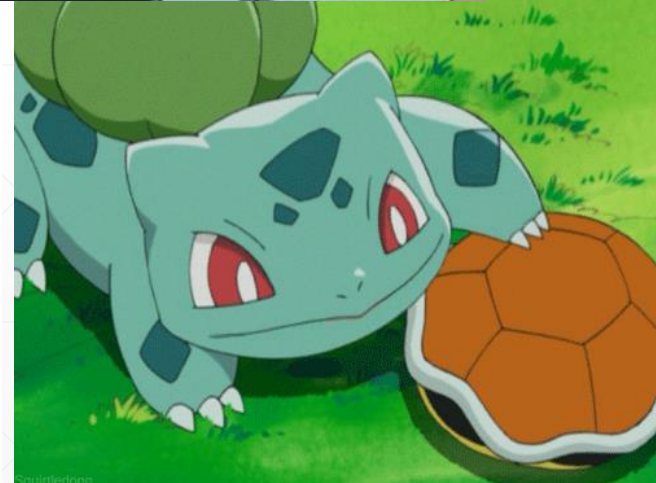# 自定义数据集实战
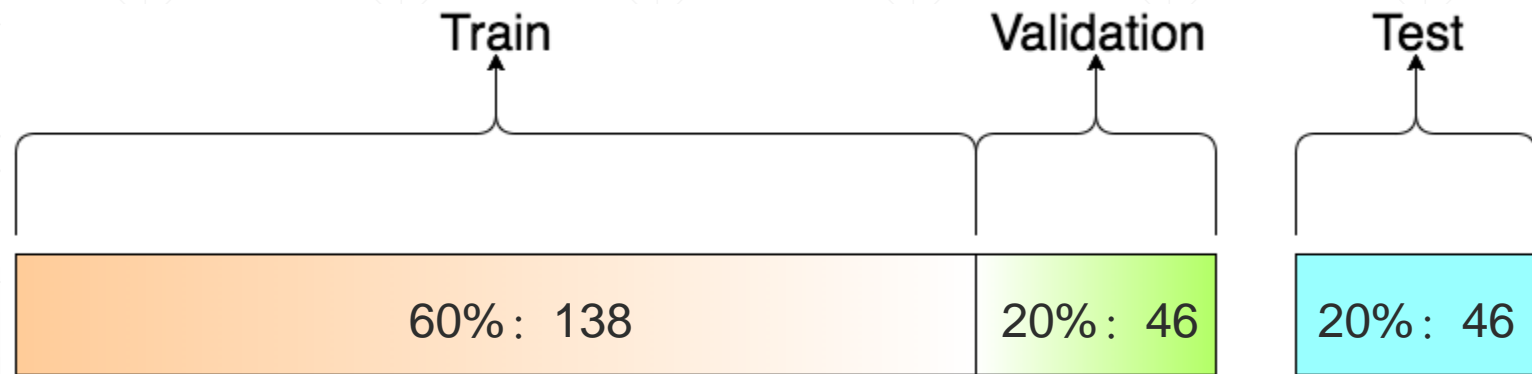
主讲：龙良曲

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-Val-Test

- Transfer Learning

Step1.Load data

- images and labels
  - $X = [1.png, 2.png, 3.png, ...]$
  - $Y = [4,9,1, ...]$

- tf.data.Dataset.from_tensor_slices((X,Y))

- .shuffle().map($func$).batch()

map($func$)

```python
def preprocess(x,y):
    x = tf.io.read_file(x)
    x = tf.image.decode_jpeg(x, channels=3)
    x = tf.image.resize(x, [224,224])

    x = tf.cast(x, dtype=tf.float32) / 255.
    y = tf.convert_to_tensor(y)

    return x, y
```

$X = [1.png, 2.png, 3.png, ...]$
$Y = [4, 9, 1, ...]$

$X = [tensor(img1), 2.png, ...]$
$Y = [tensor(4), 9, 1, ...]$

Custom Dataset

```python
images, labels, table = load_pokemon('pokemon', 'train')
print('images', len(images), images)
print('labels', len(labels), labels)

db = tf.data.Dataset.from_tensor_slices((images,labels))

db = tf.data.Dataset.from_tensor_slices((images, labels))
db = db.shuffle(1000).map(preprocess).batch(32)
```
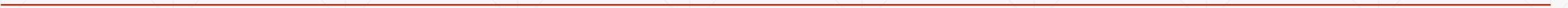
Preprocessing

- Read and Resize
  - 224x224 for ResNet

- Data Augmentation
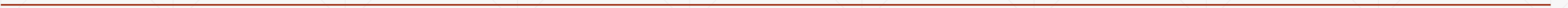 - Rotate/Flip
  - Crop

- Normalize
  - Mean, std

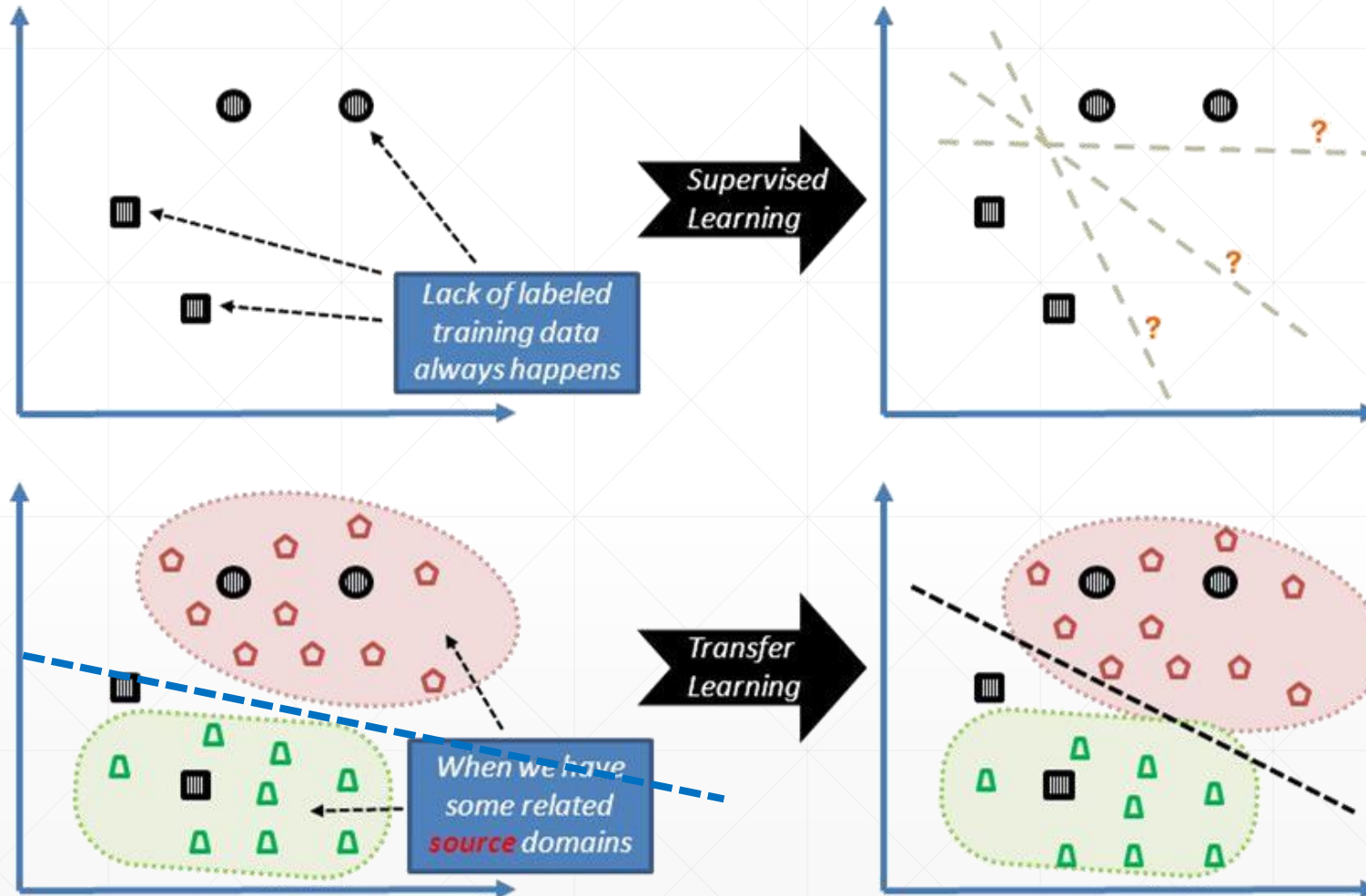# Step2.build model

- Inherit from Model

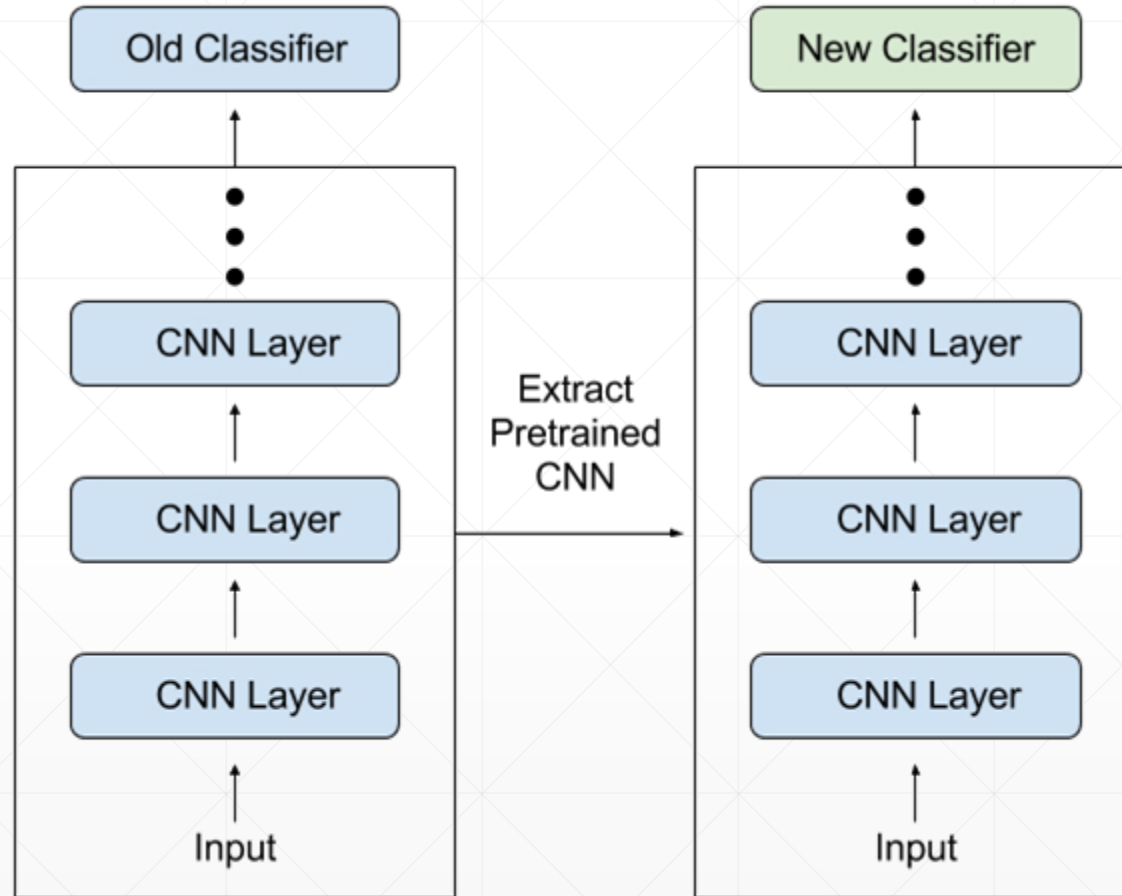- Define forward graph

Step3.Train and Test

- Early stopping

Step4.Transfer Learning

Transfer learned representation

In Conclusion

- Load custom data

- Train from scratch

- Transfer learning

下一课时

Thank You.