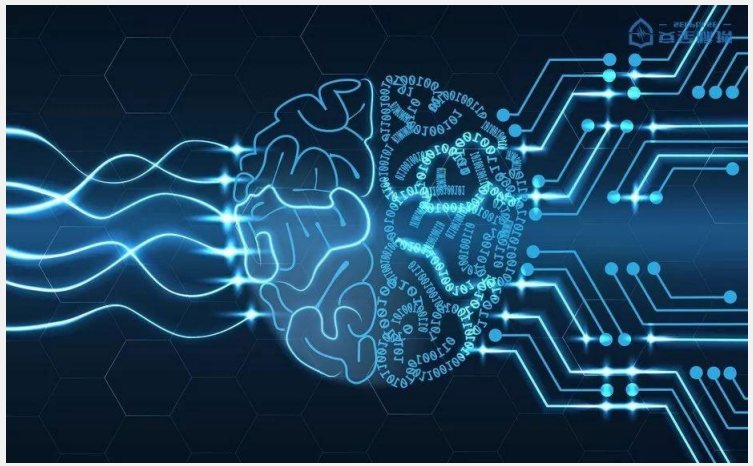


机器学习

Machine Learning

授课老师：谭毅华
电　　话：13886021197
办 公 室：科技楼1102
邮　　箱：*yhtan@hust.edu.cn*

第十二章、迁移学习

目录
CONTENTS

1.迁移学习问题

- 《论语·为政》：温故而知新，可以为师矣

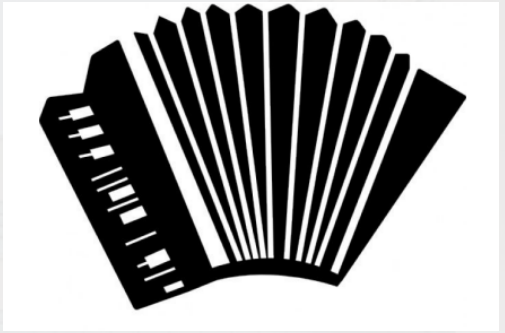
- 《庄子·天运》：故西施病心而颦其里，其里之丑人见而美之，归亦捧心而颦其里。其里之富人见之，坚闭门而不出；贫人见之，挈妻子而去之走。彼知颦美，而不知颦之所以美。

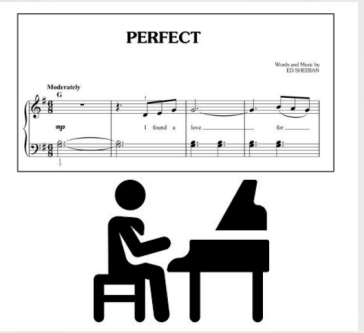
- 旧的知识可以提炼升华，迁移到新知识的学习
- 东施效颦以失败告终，表明旧知识需正确使用才能发挥作用
- 关键在于不同事物之间的相关性

迁移学习问题：如何有效地利用事物之间的相关性，来帮助我们解决新问题、学习新能力

1.迁移学习

- 概念：指利用数据、任务、或模型之间的相似性，将在旧领域学习过的模型，应用于新领域的一种学习过程。
  - **球类运动相似：乒乓球高手可很快学会打网球**
  - **乐器相通：手风琴和钢琴也有相通的地方**





- 动机：迁移学习可以发挥什么作用

- **减少对标记数据的依赖**

社交微博翻译任务，可借助书面语翻译的模型减少对标记数据的依赖

- **降低对机器算力的需求**

降低模型训练的复杂度，无需高性能服务器集群也可完成模型训练

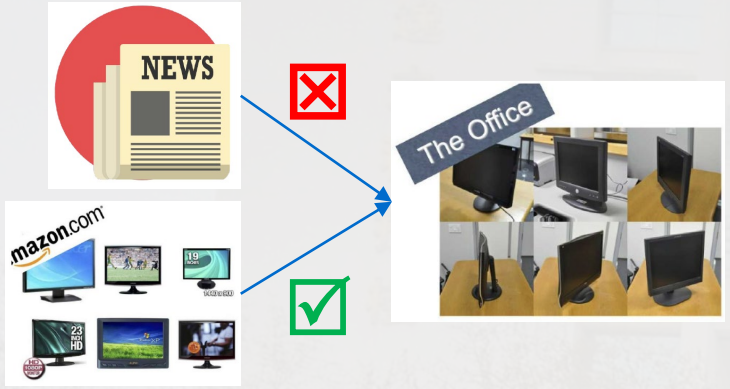
- **使模型适配个性化任务**

借助网上搜索到的显示器图片，辅助训练办公场景下显示器的检测模型

1.迁移学习

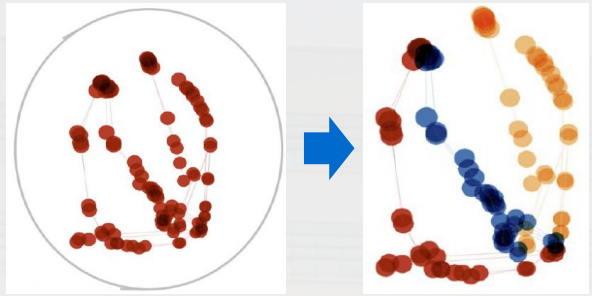
迁移学习的三个基本问题

- **何时迁移**



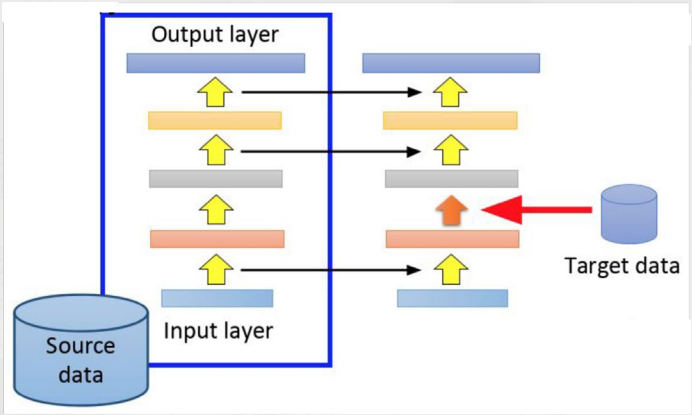
原问题和目标问题具有相关性

- **何处迁移**



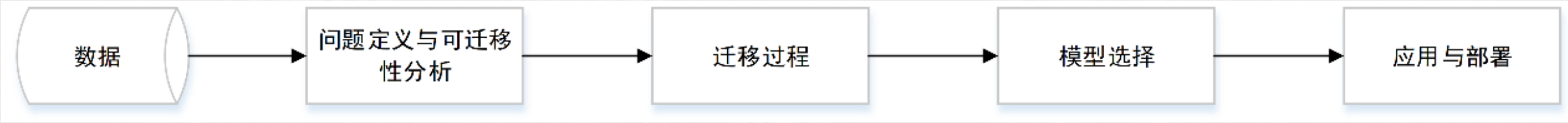
要迁移的知识往往来源于相似的模式或特征

- **如何迁移**



建立原问题和目标问题的关联

迁移学习的一般过程



2.域适应

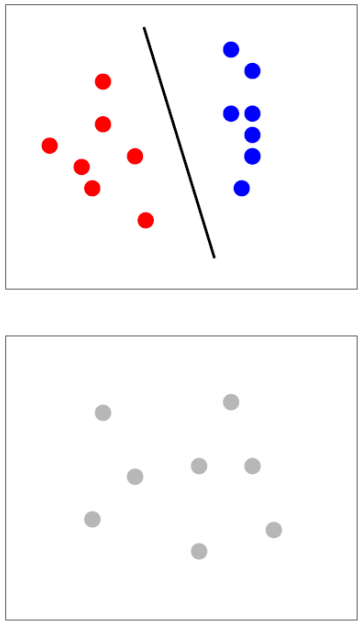
- 迁移学习一般可形式化为**域适应**问题

- **定义：以分类问题为例，域适应的定义如下：**

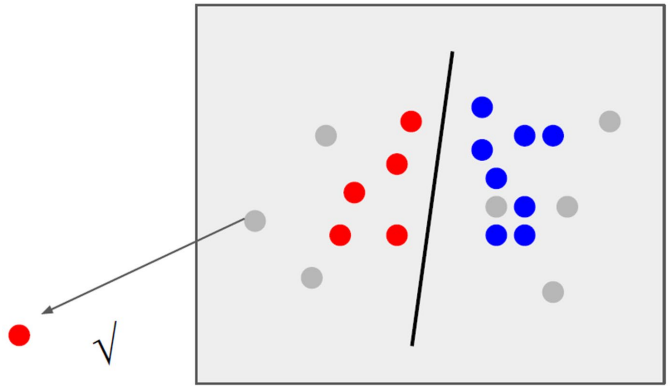
给定源域数据集$D_s = \{x_i, y_i\}_{i=1}^{n}$（包括数据和标签），以及目标域数据集$D_t = \{x_j\}_{j=n+1}^{n+m}$（往往不含标签），利用$D_s$和$D_t$学习一个分类器$f: x_t \to y_t$，该分类器输入目标域的数据，输出对应的分类结果。

源域数据集$D_s = \{x_i, y_i\}_{i=1}^{n}$
（含数据和标签）

目标域数据集$D_t = \{x_j\}_{j=n+1}^{n+m}$
（含数据，不含标签）



域适应方法

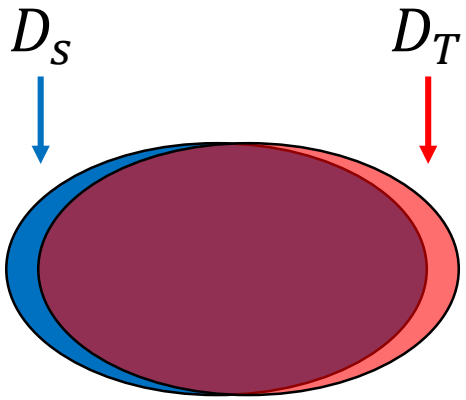


可分类目标域数据的分类器f

2.域适应

现有域适应方法大致分为三类

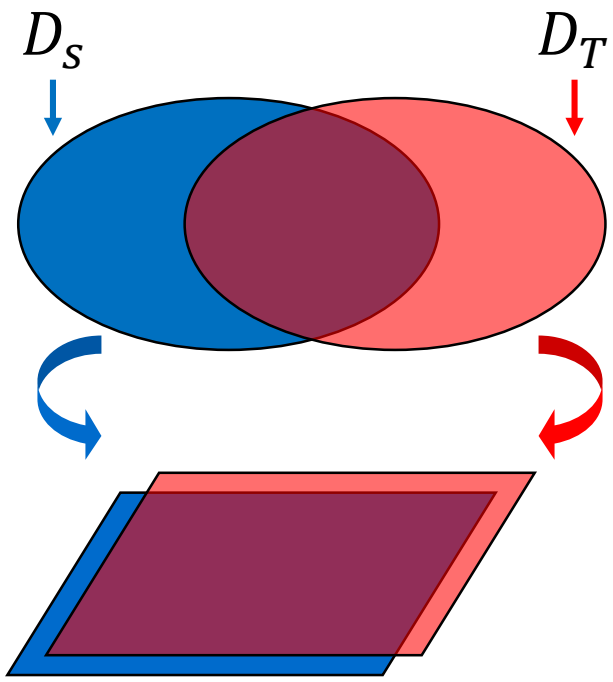
● 基于实例的域适应



假设源域和目标域数据重叠度较高，在模型训练时重点关注与目标域相似的源域样本

$$\min \frac{1}{n}\sum_{i=1}^{n} w_i L(\Phi(x_i^s), y_i^s, \theta)$$

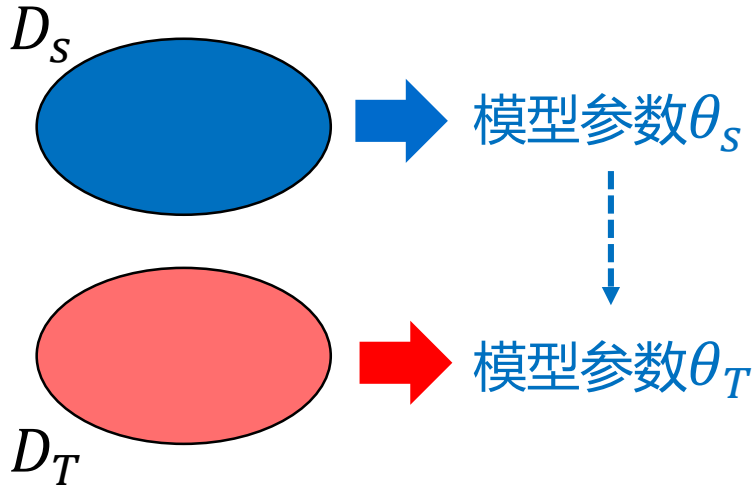
● 基于特征的域适应（目前最常用的方法）



将源域样本和目标域样本映射到同一个特征空间，使样本在这个特征空间能够“对齐”

$$\min \frac{1}{n}\sum_{i=1}^{n} L(\Phi(x_i^s), y_i^s, \theta)$$

● 基于模型参数的域适应



先利用源域数据训练模型的初始参数，再通过参数迁移使得模型更好地在目标域上工作

$$\min \frac{1}{n}\sum_{i=1}^{n} L(\Phi(x_i^s), y_i^s, \theta')$$

3.最大均值差异法

- 基于最大均值差异的域适应方法是一种经典算法
  - **以分类问题为例，域适应的典型优化目标为：**

$$f^* = \arg\min_f \frac{1}{N_s} \sum_{i=1}^{N_s} \ell(f(x_i), y_i) + \lambda D(D_s, D_t)$$

N_s: 源域样本数目

f: 待训练分类器

x_i: 源域输入样本

y_i: 源域样本类别标签

l: 分类损失

D_s: 源域样本集合

D_t: 目标域样本集合

D: 源域和目标域之间的距离度量

该优化目标的含义：最小化源域和目标域的特征距离的同时，最小化源域样本的分类损失

关键：定义源域和目标域的特征距离度量D，常选用最大均值差异

3.最大均值差异法

- 最大均值差异：可度量特征空间中两个分布的距离
  - **最大均值差异的一般形式：**

$$\mathrm{MMD}^2(A,B)=\left\|\sum_{i=1}^{n1}\phi(a_i)-\sum_{j=1}^{n2}\phi(b_j)\right\|^2$$

A, B: 两个样本集合　a_i: 集合A中的样本　b_j: 集合B中的样本

ϕ: 特征映射函数　$\phi(b_j)$: 样本b_j的映射特征

$\phi(a_i)$: 样本a_i的映射特征

直观解释：计算两个集合中样本的特征均值的距离

3.最大均值差异法

- 使用最大均值差异度量源域和目标域的距离 $f^* = \arg\min_f \frac{1}{N_s}\sum_{i=1}^{N_s} \ell(f(x_i), y_i) + \lambda D(D_s, D_t)$

  - **源域和目标域的距离，可分解为边缘分布和条件分布的最大均值差异：**

$$D(D_s, D_t) \approx (1-\mu) D(P_s(x), P_t(x)) + \mu D(P_s(y \mid x), P_t(y \mid x))$$
$$= (1-\mu) MMD(P_s(x), P_t(x)) + \mu MMD(P_s(y \mid x), P_t(y \mid x))$$

$P_s(x)$, $P_t(x)$: 源域和目标域中样本的边缘分布， $P_s(y|x)$, $P_t(y|x)$: 源域和目标域中样本的条件分布

  - **边缘分布的最大均值差异可表示为：**

$$MMD(P_s(x), P_t(x)) = \left\| \frac{1}{N_s}\sum_{i=1}^{N_s} A^T x_i - \frac{1}{N_t}\sum_{j=1}^{N_t} A^T x_j \right\|_{\mathrm{H}}^2$$

A: 将样本映射为特征的变换矩阵，是待优化的变量

  - **条件分布的最大均值差异可近似表示为：**

$$MMD(P_s(y \mid x), P_t(y \mid x)) \approx \sum_{c=1}^{C} \left\| \frac{1}{N_s^{(C)}} \sum_{x_i \in D_s^{(c)}} A^T x_i - \frac{1}{N_t^{(C)}} \sum_{x_i \in D_t^{(c)}} A^T x_j \right\|_{\mathrm{H}}^2$$

(c): 标识第c个类别的对应变量
目标域的(c)信息为估计的伪标签

3.最大均值差异法

优化目标的化简

- **再考虑特征映射后的散度最大化，即方差最大：** $\max(A^T X)H(A^T X)^T$ $H = I - (1/n)1$
- **最小化源域和目标域的均值差异及散度约束，其优化目标可化简为：**

$$\min D(D_s, D_t) \Rightarrow \min \frac{tr(A^T XMX^T A)}{tr(A^T XHX^T A)} \Rightarrow \min tr(A^T XMX^T A) + \lambda \|A\|_F^2, \quad s.t. A^T XHX^T A = I.$$

$$\max(A^T X)H(A^T X)^T$$

X: 由源域和目标源域样本拼接成的矩阵

M: 称为MMD矩阵，由边缘和条件MMD矩阵线性组合得到

H: 称为中心矩阵，由单位矩阵和常数矩阵相减得到

- **使用拉格朗日法进行求解上式，求解所得矩阵A即用于对齐源域和目标域的特征映射矩阵**
- **以$(A^T X_s, Y_s)$训练分类器f_i,再对$A^T X_t$进行测试**

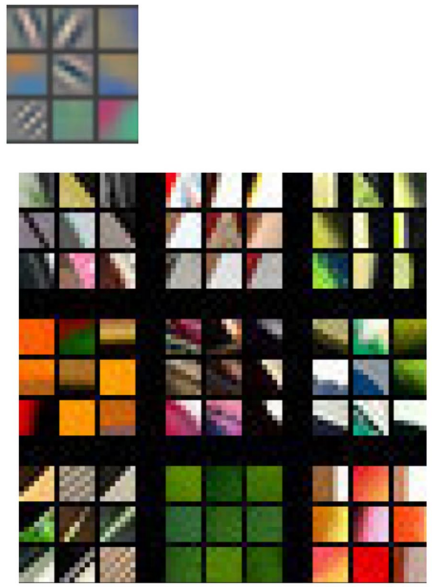
Ref: Wen Zhang, Dongrui Wu. "Discriminative Joint Probability Maximum Mean Discrepancy (DJP-MMD) for Domain Adaptation", Int'l Joint Conf. on Neural Networks (IJCNN), Glasgow, UK, 2020.

https://blog.sciencenet.cn/home.php?mod=space&uid=3418535&do=blog&id=1227915

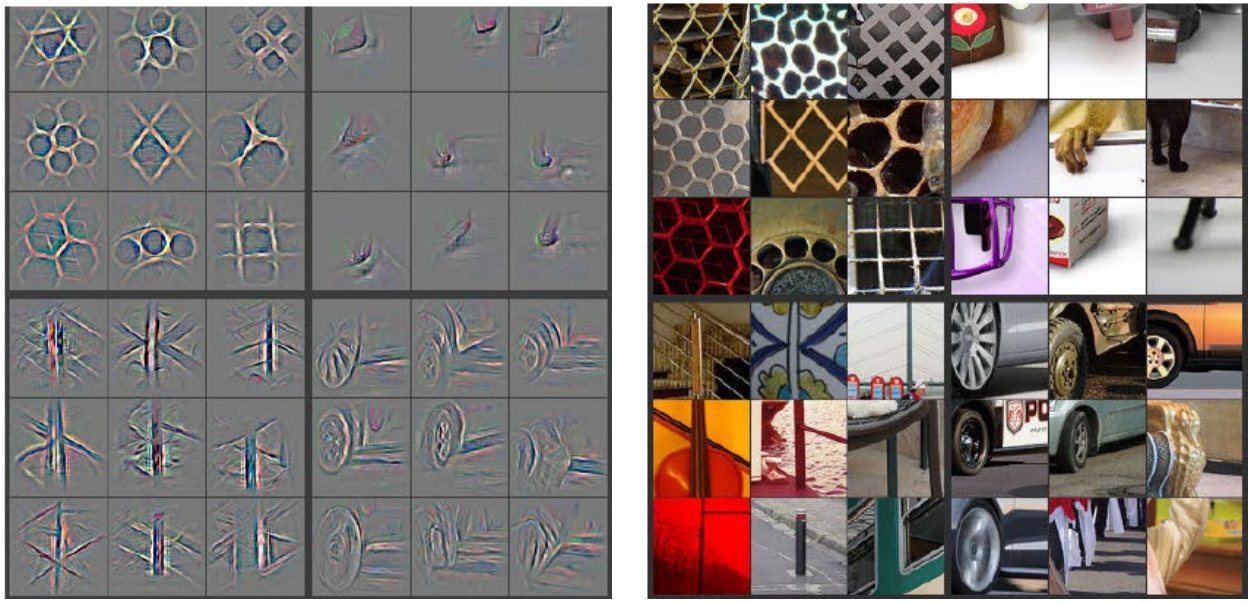
4.深度学习中的迁移学习

深度网络的可迁移性

- **深度网络在前几层关注边缘、纹理等低层信息，后续层关注整体形状、结构等高层信息**



网络第一层特征的可视化



网络第三层特征的可视化

不同任务往往都依赖于边缘、纹理等信息，
因此在源域上学习的一部分特征，在目标域上往往也有效。

4.深度学习中的迁移学习

- 第一种方式：基于Finetune（微调）的迁移学习

开源大规模数据集

从头训练

深度学习
预训练模型

特定任务的小规模数据集

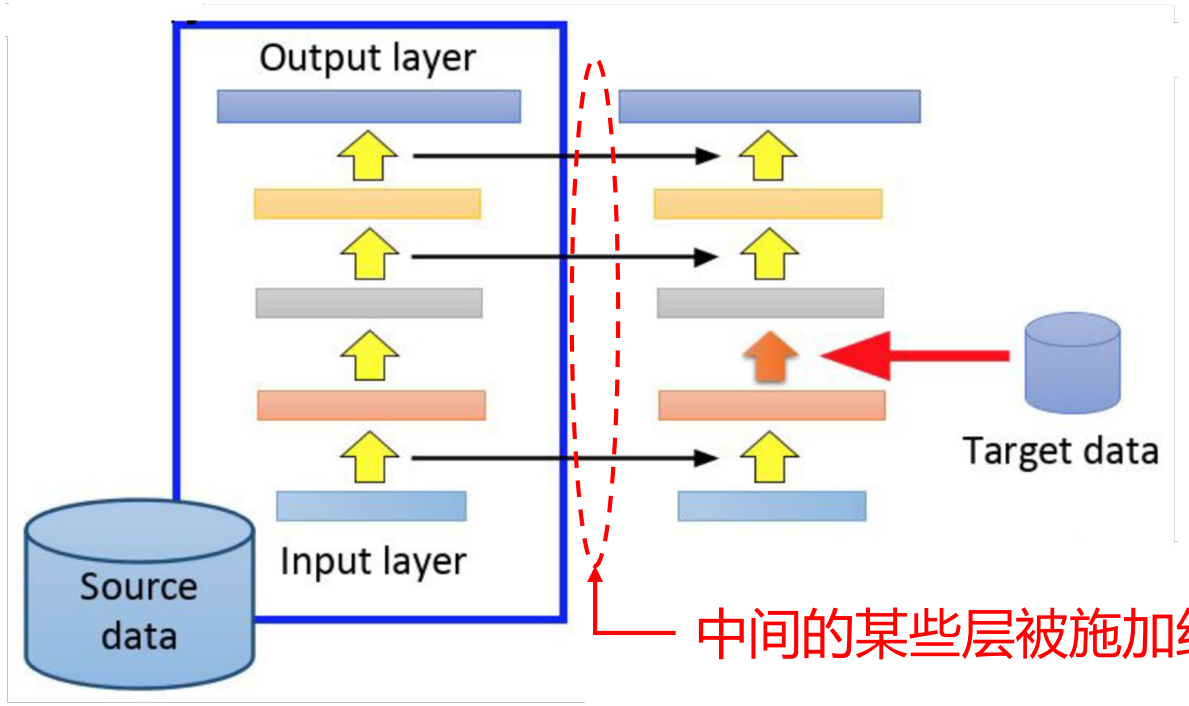
Finetune（微调）

面向特定任务的
深度学习模型

4.深度学习中的迁移学习

第二种方式：基于自适应层的迁移学习

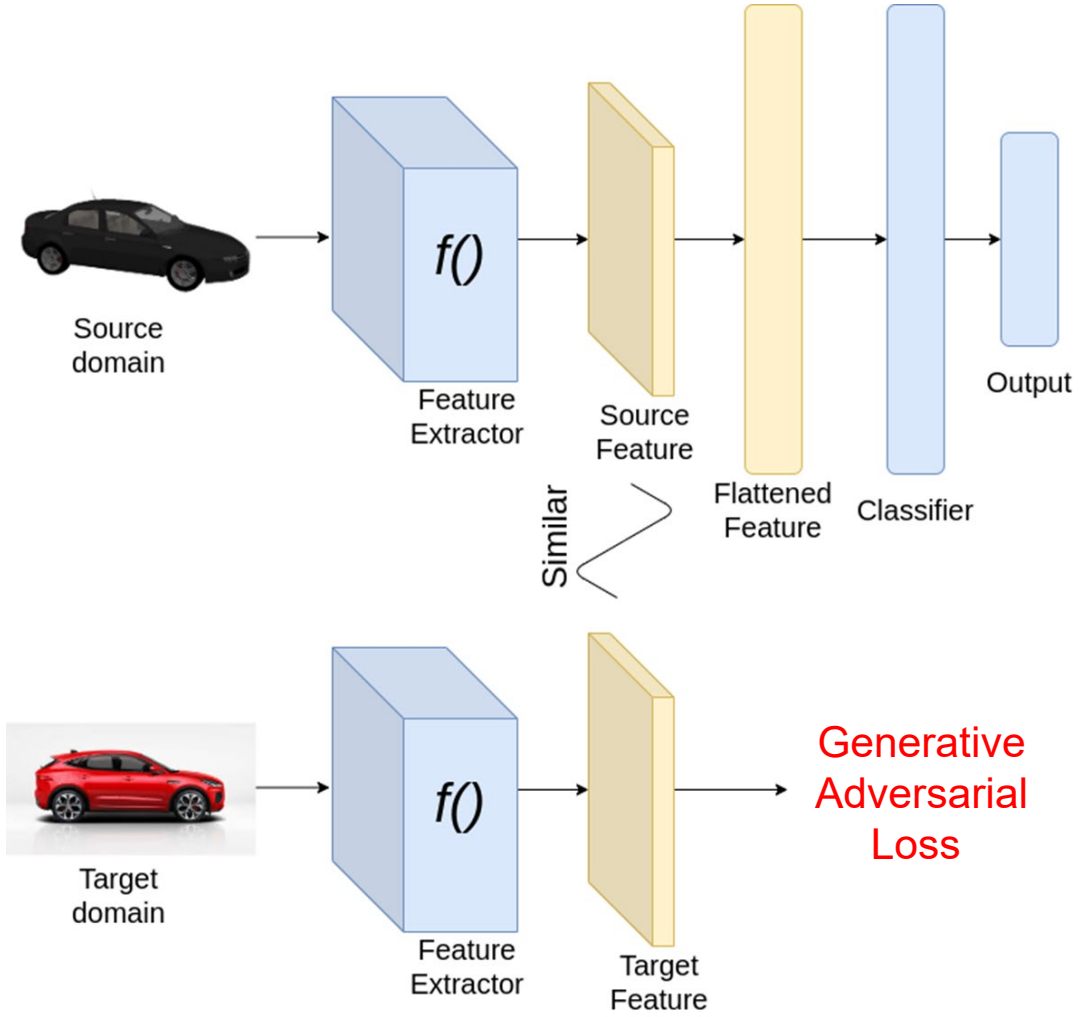
- **源域和目标域之间差异较大时，此类方法往往优于基于Finetune的方法**



核心思想是使得源域和目标域的特征分布更加接近

4.深度学习中的迁移学习

深度对抗网络迁移：第二种方式的代表性实现



目标：源域样本被正确分类

最终目标：目标域样本被正确分类

目标：源域特征和目标域特征十分相似

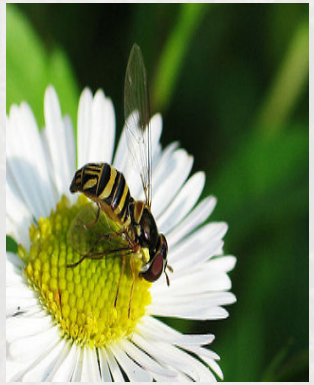
$$\ell=\ell_c(D_s,y_s)+\lambda\ell_d(D_s,D_t)$$

5.迁移学习实例

花卉种类识别：基于Finetune（微调）的迁移学习

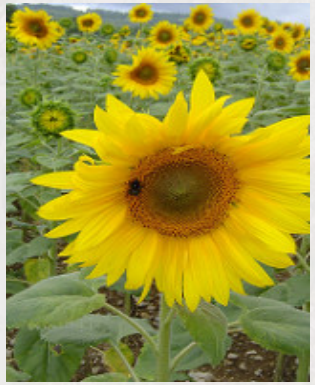
- **数据集**

花卉名称	训练集（张）	验证集（张）
雏菊	1699	200
蒲公英	2494	200
玫瑰花	1723	200
向日葵	1897	200
郁金香	2197	200


雏菊


蒲公英


玫瑰

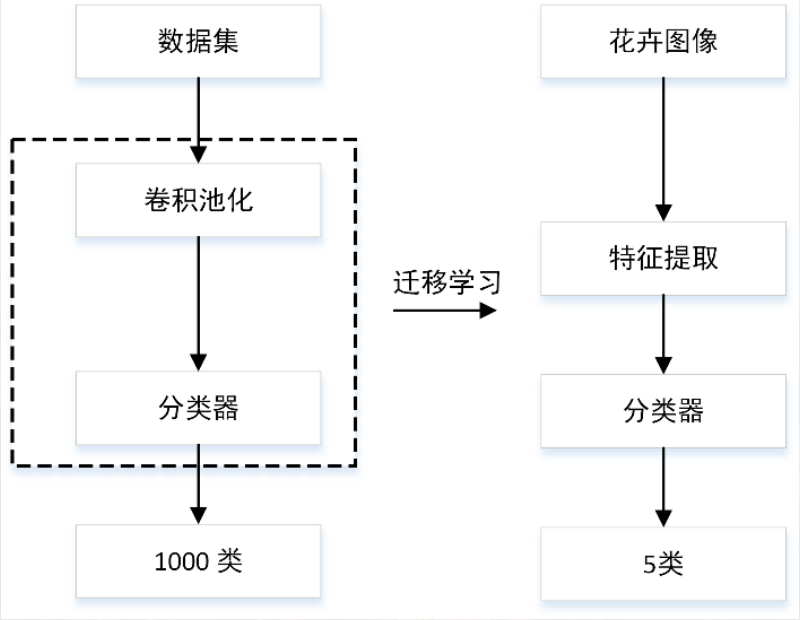

向日葵


郁金香

5.迁移学习实例

花卉种类识别：基于Finetune（微调）的迁移学习

迁移学习过程



ImageNet上预训练后，在花卉图像上微调

实验结果

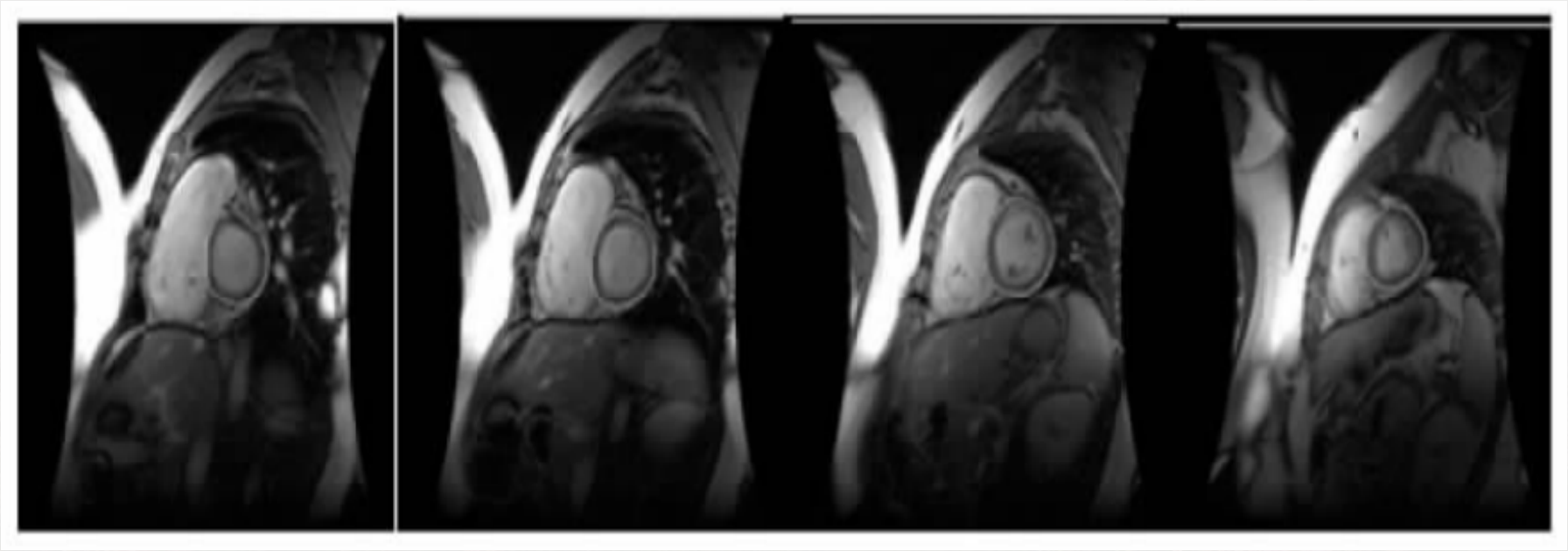
方法	分类识别率(%)				
	雏菊	蒲公英	玫瑰花	向日葵	郁金香
SVM	50.61	47.58	45.69	54.29	57.27
CNN	82.52	81.25	84.19	80.11	81.24
深度迁移学习	94.25	95.14	92.21	93.17	93.89

第一行：训练支持向量机进行识别
第二行：从头训练CNN进行识别
第三行：迁移学习的方式训练CNN进行识别

实验结论：深度迁移学习有效提升了花卉识别精度

5.迁移学习实例

- 医学图像分割：基于逐层微调和生成对抗损失的迁移学习
  - **样本示例**

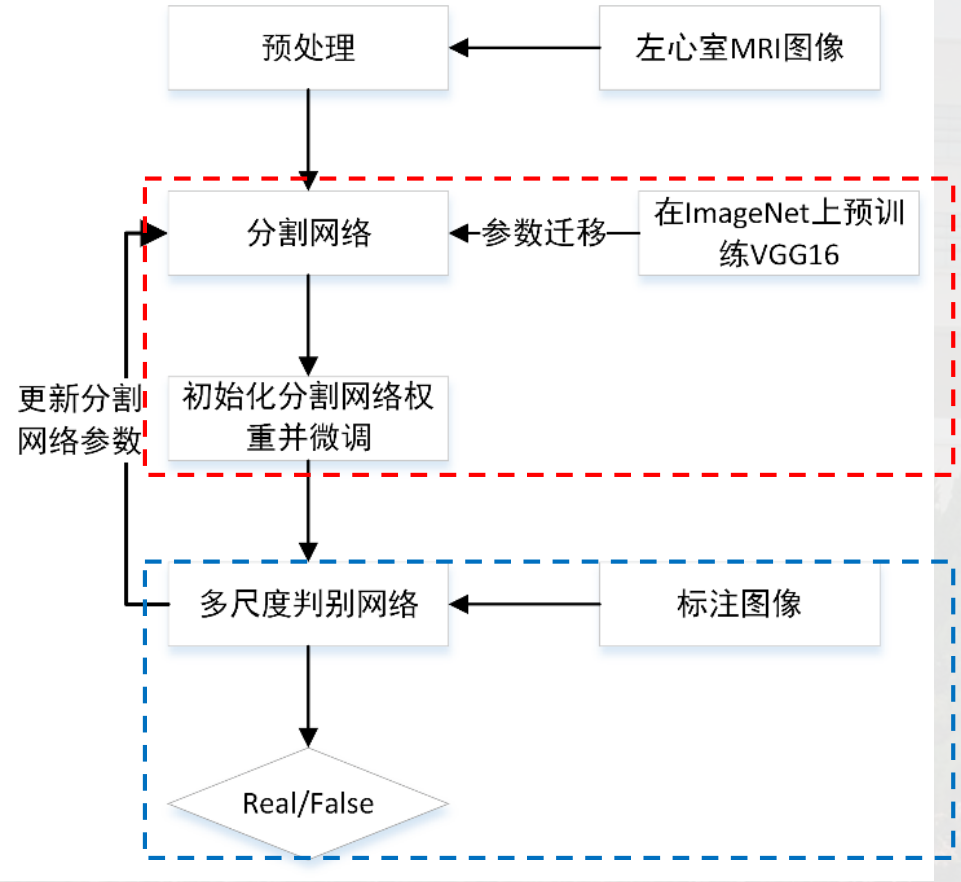


任务：自动分割出心脏的不同部位

5.迁移学习实例

医学图像分割：基于逐层微调和生成对抗损失的迁移学习

分割网络整体架构图



复用参数并微调，迁移大规模数据上习得的知识

$$L_{mae}(F(x),y)=\frac{1}{N}\sum_{n=1}^{N}\|F(x_i)-Y_i\|$$

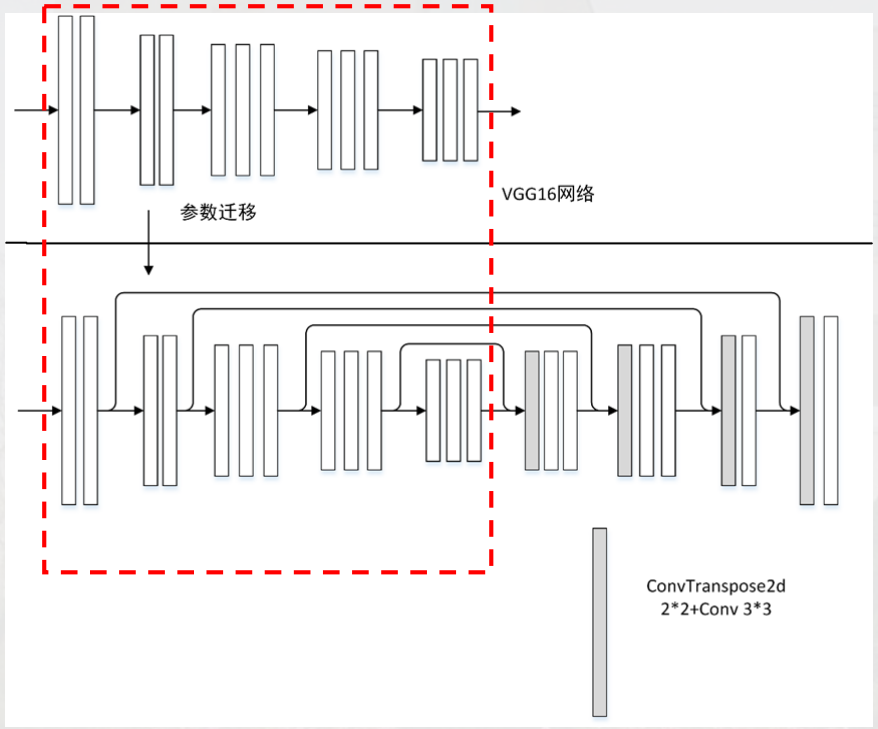
引入生成对抗损失，提高分割结果的真实性

$$\min_{G}\max_{D}V(G,D)+\beta L_{mae}(F(x),y)$$

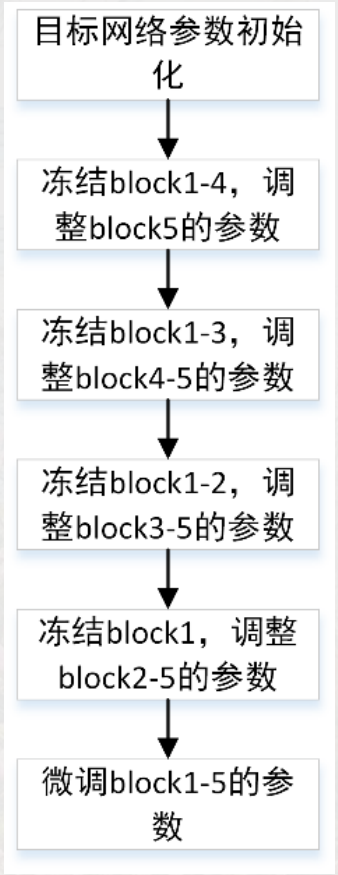
5.迁移学习实例

医学图像分割：基于逐层微调和生成对抗损失的迁移学习

基于迁移学习的分割网络结构示意图



逐层微调流程图



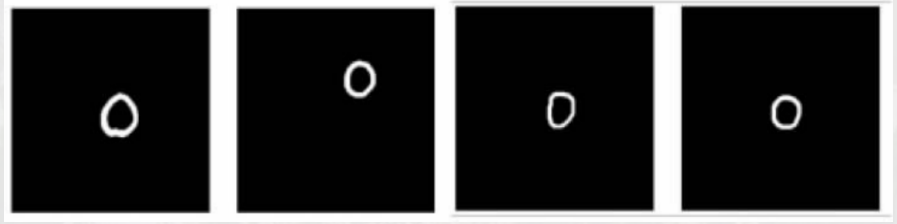
分割网络的编码器部分，复用VGG16网络中的对应层。
VGG16网络使用大规模数据集ImageNet进行预训练。

5.迁移学习实例

医学图像分割：基于逐层微调和生成对抗损失的迁移学习

实验结果

专家手动分割



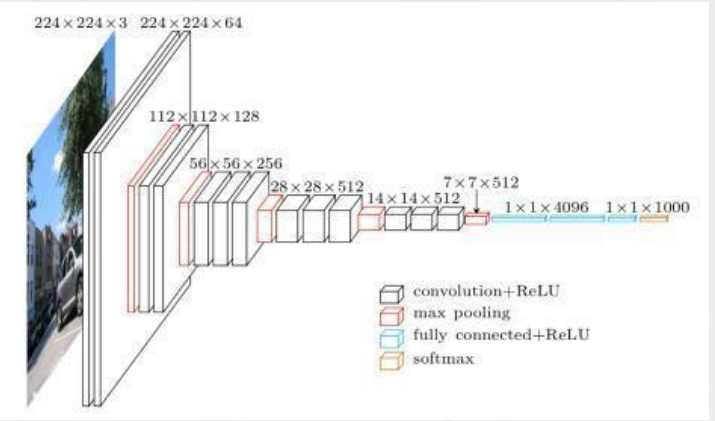
算法分割

方法	Dice		Jaccard		Sensitivity	
	内膜	外膜	内膜	外膜	内膜	外膜
Nasr(2018)	~	0.8724	~	~	~	0.8769
Nasr(2018)	0.9258	0.9606	0.8692	0.9037	0.9224	0.9583
U-net	0.8829	0.9292	0.8536	0.8843	0.8910	0.9314
TLBSN	0.9307	0.9623	0.8764	0.9281	0.9232	0.9503
迁移学习对抗网络	0.9399	0.9697	0.8968	0.9415	0.9363	0.9686

实验结论：迁移学习对抗网络有效提高了分割精度，验证了迁移学习的有效性。

5.迁移学习实例

- 代码讲解：以花卉图像识别为例
- **核心代码：加载预训练的VGG16模型，并调整其结构为二分类模型**




```
model = applications.VGG16(weights="imagenet", include_top=False,
                           input_shape=(img_width, img_height, 3))
for layer in model.layers[:5]:
    layer.trainable = False
x = model.output
x = Flatten()(x)
x = Dense(1024, activation="relu")(x)
x = Dropout(0.5)(x)
x = Dense(1024, activation="relu")(x)
predictions = Dense(2, activation="softmax")(x)
model_final = Model(inputs=model.input, output=predictions)
```

← 加载预训练VGG16模型

← 冻结VGG16模型的前5层参数，这部分参数将不会被更新

← 在VGG16基础模型上追加两个全连接层，这两层的参数为随机初始化

← 最后追加一个输出维度为2的全连接层，对应二分类结果

• 本章小结

- 建立迁移学习和域适应的概念
- 域适应的三种主要方式：基于实例、特征或模型参数的域适应
- 基于最大均值差异的域适应方法的目标函数和优化思路
- 深度学习中的迁移学习主要包括两类方法：基于微调或自适应层的迁移学习
- 深度网络迁移学习的一般流程：预训练、引入自适应层的约束（可选）、迁移训练