北京大学
PEKING UNIVERSITY

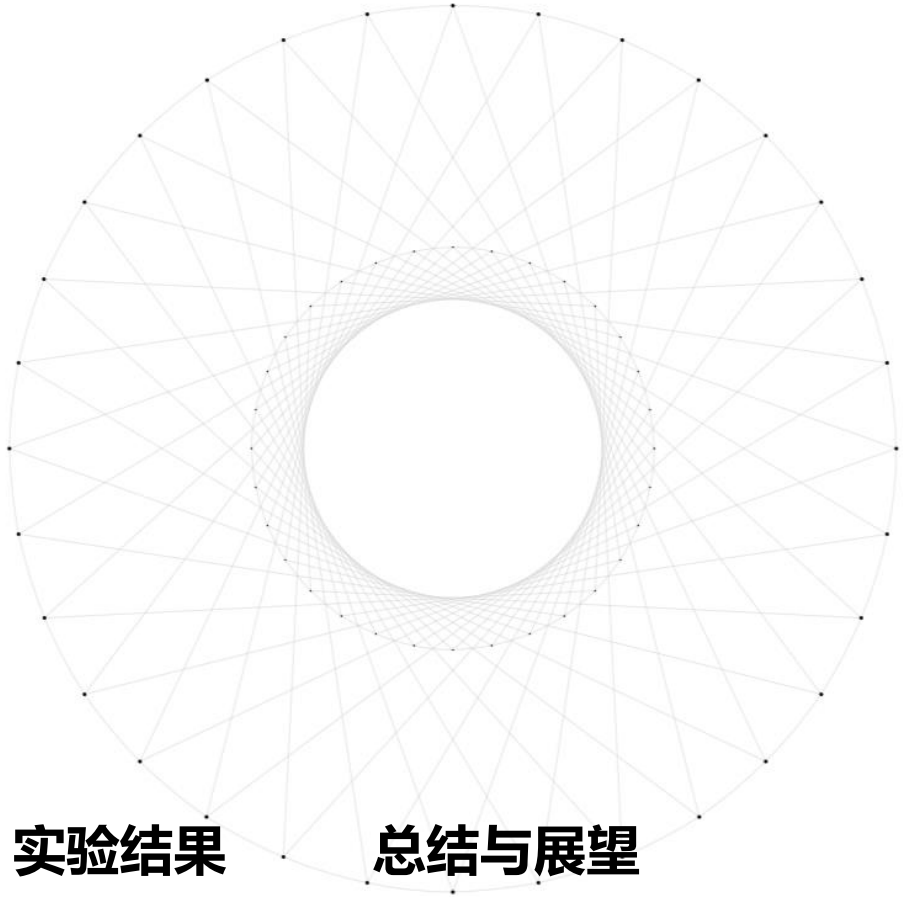
# 图像分类的非监督领域迁移

组员：陈亦弘 张鉴心

学院：前沿交叉学科研究院

北京大学
PEKING UNIVERSITY

# 目录
CONTENTS





北京大学
PEKING UNIVERSITY

# 背景介绍

## CHAPTER ONE

北京大学
PEKING UNIVERSITY

## 背景介绍

- 任务的目标是把一个在有标注的数据集上训练出来的分类模型迁移到另一个没有标注的数据集上。

- 深度学习在拥有大量标注的数据的情况下可以获得很好的表现，但是随之而来的问题是有标注的数据可能需要大量的人力和财力。所以如果能够把一个在有大量标注的数据集上训练出来的模型迁移到另一个具有相同标签、但是图像分布不同的数据集中，就可以极大地减少资源的使用。

北京大学
PEKING UNIVERSITY

## 背景介绍

生成对抗学习中的生成器通过将源域分布不断改进，使源域分布近似于目标与分布，从而让域判别器无法识别出数据来自哪一域。域适应是迁移学习的一部分，迁移学习则是想利用类似于目标分布的源分布，从中迁移出对目标任务分类有益的“知识”，或者是将源分布与目标分布映射到“共同特征空间”，完成对目标任务的无监督/半监督/少样本学习等。对抗学习与迁移学习的融合是当前迁移学习领域中的一个热点，一些方法通过将对抗学习用于无监督域适应，确实减少了源域和目标域之间的差异，并提高了泛化能力，但是其中也存在一些改进的地方，比如CoGAN对源域和目标域之间不太相似的情况下表现的不太满意。

北京大学
PEKING UNIVERSITY

# 现有方法

CHAPTER TWO

北京大学
PEKING UNIVERSITY

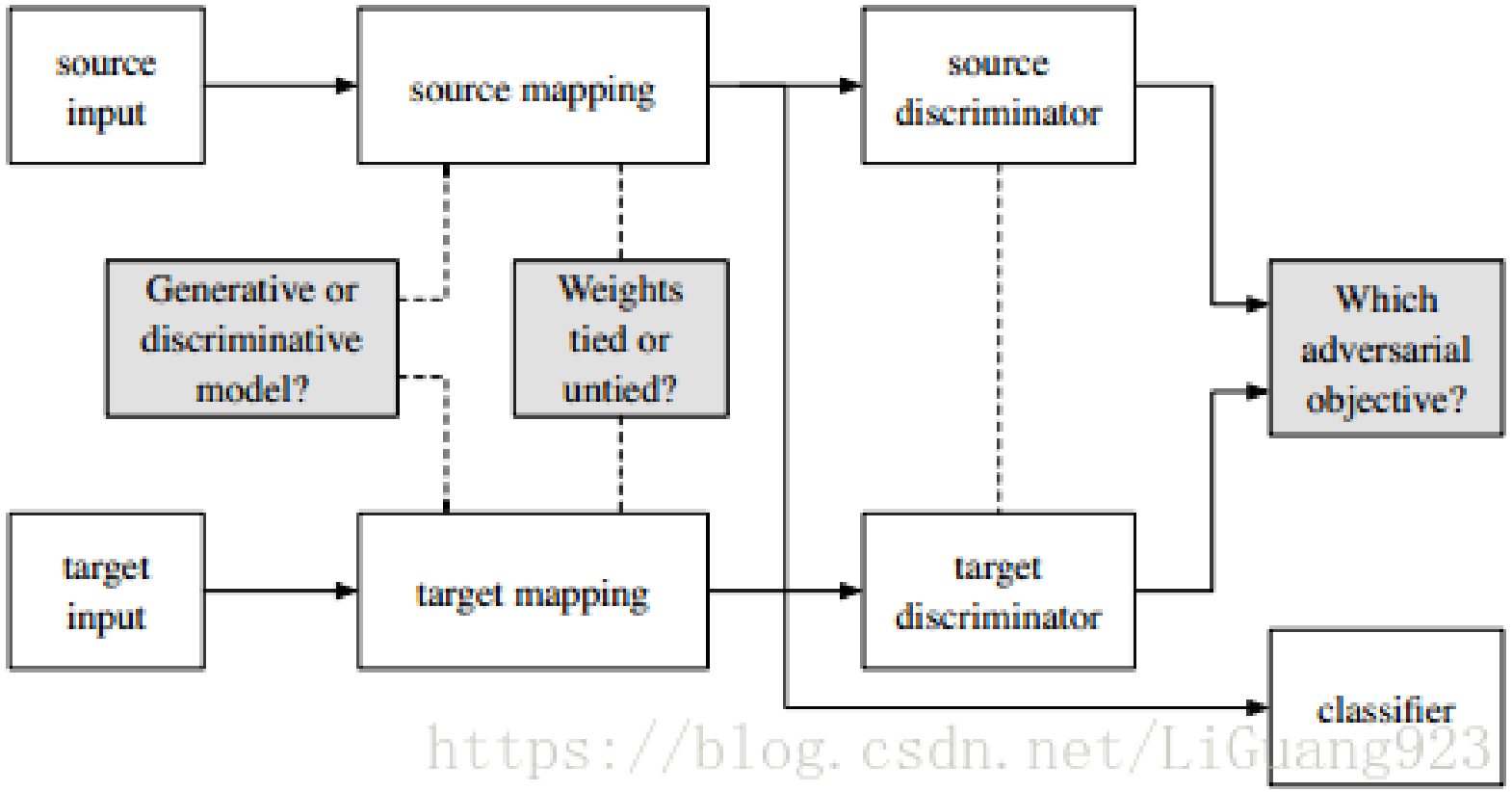
## 现有方法

○ 在论文Adversarial Discriminative Domain Adaption中作者认为，之前不同的对抗自适应算法主要有以下三点区别：

- 生成式模型还是判别式模型
- 不同域的映射结构是否共享权重
- 使用何种对抗损失函数

北京大学
PEKING UNIVERSITY

## 现有方法



北京大学
PEKING UNIVERSITY

## 现有方法

- 区别1：生成模式还是判别模式

生成式模型用随机噪声作为输入，在图像空间产生样本。一般会使用判别器的中间层特征来训练一个任务相关的分类器。判别模型则会直接将图片映射到特征空间，然后输入到分类器中进行训练。

- 区别2：不同域的映射结构是否共享权值

很多之前的对抗自适应方法都采用源域和目标域的映射结构共享权重的方式。这样做可以减少模型的参数，同时保证这样的映射至少在目标域上是有判别力的。但是这样同一个网络需要处理来自两个不同域的图片，这样可能会在优化过程中出现病态条件。还有一些方法只对一部分层进行权重的共享，如CoGAN。当然也可以使源域和目标域的映射结构完全不共享参数。

北京大学
PEKING UNIVERSITY

## 现有方法

○ 区别3：使用何种对抗损失函数

- Minmax loss：

$$\mathcal{L}_{\text{adv}_M} = -\mathcal{L}_{\text{adv}_D}$$

- GAN loss：

$$\mathcal{L}_{\text{adv}_M}(\mathbf{X}_s, \mathbf{X}_t, D) = -\mathbb{E}_{\mathbf{x}_t \sim \mathbf{X}_t}[\log D(M_t(\mathbf{x}_t))]$$

- Domain confusion loss：

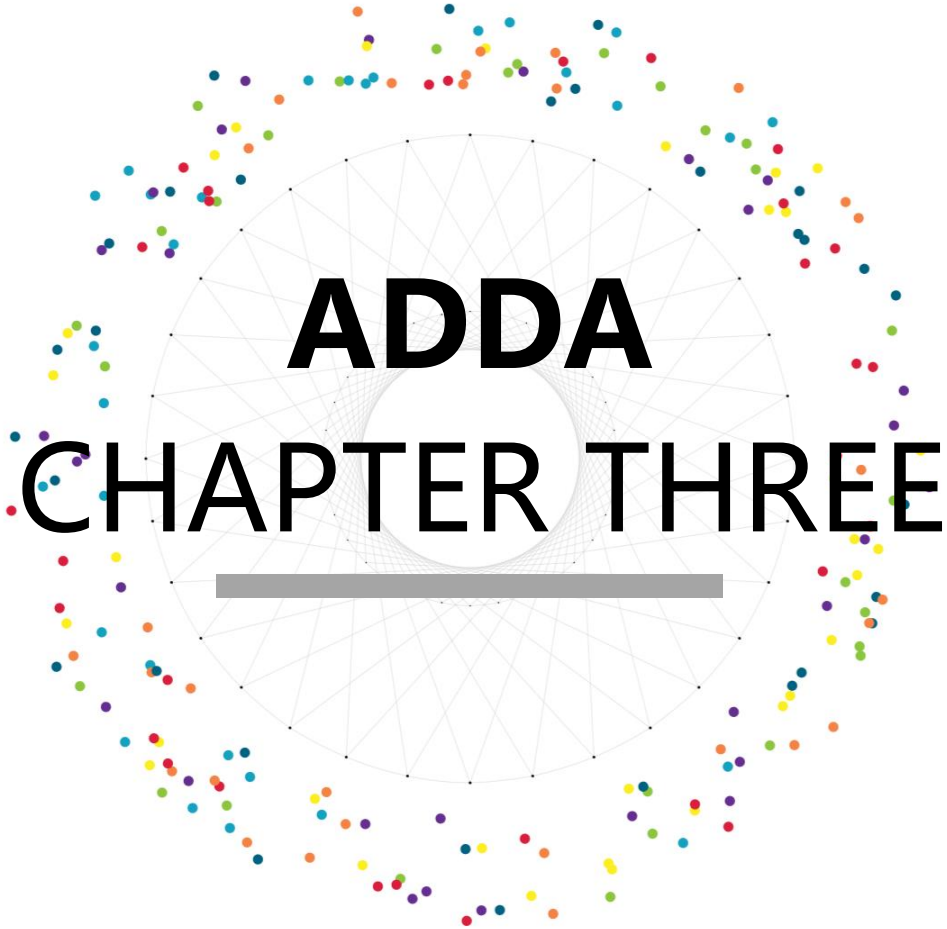
$$\mathcal{L}_{\text{adv}_M}(\mathbf{X}_s, \mathbf{X}_t, D) = -\sum_{d \in \{s,t\}} \mathbb{E}_{\mathbf{x}_d \sim \mathbf{X}_d}\left[\frac{1}{2}\log D(M_d(\mathbf{x}_d)) + \frac{1}{2}\log(1 - D(M_d(\mathbf{x}_d)))\right]$$

北京大学
PEKING UNIVERSITY

## 现有方法

- Deep Domain Confusion: Maximizing for Domain Invariance是基于特征的变换，采用maximum mean discrepancy，source domain和target domain之间的参数完全共享。
- Unsupervised Domain Adaptation by Backpropagation是基于特征变换，采用GAN loss。在特征提取之后，在域分类器之前加入了一个梯度反转层。
- Collaborative and Adversarial Network for Unsupervised domain adaptation使用多个判别器，同时从特征提取器的底层提取域相关信息和从特征提取器的高层提取无关信息。

北京大学
PEKING UNIVERSITY



# ADDA

## CHAPTER THREE

北京大学
PEKING UNIVERSITY

# ADDA

## 模型说明

- 首先，作者使用判别模型。因为作者认为用于生成样本的大量参数与要执行的判别任务无关。

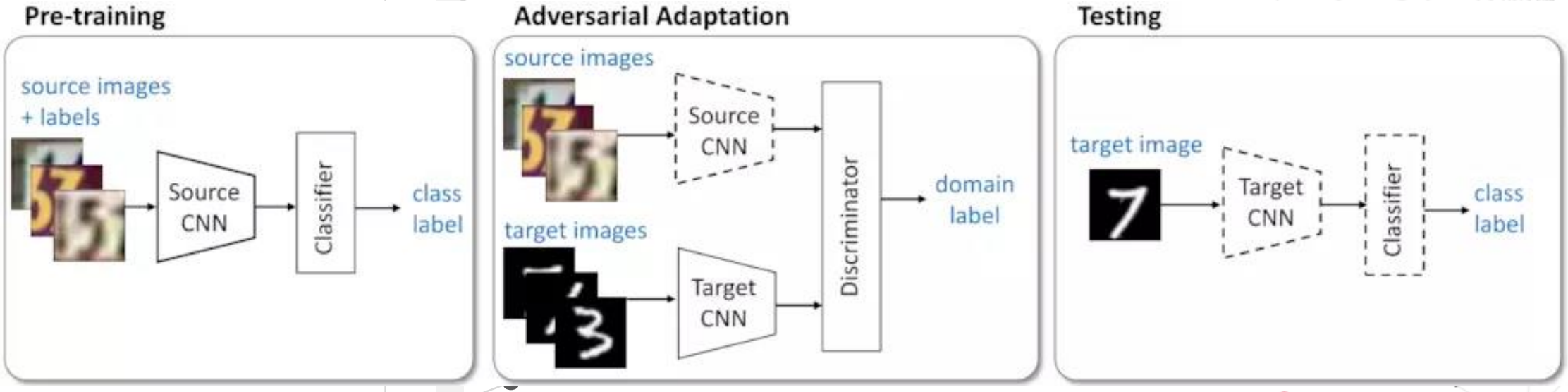
- 其次，作者使用独立的源域和目标域映射网络，两部分不共享参数。这是一个更灵活的设计，可以让映射网络学习到更多的特定领域特征。作者用预训练的源域映射网络权重初始化目标域映射网路。

- 最后，作者用GAN loss作为映射网络的对抗损失。

北京大学
PEKING UNIVERSITY

# ADDA

## 模型框架



北京大学
PEKING UNIVERSITY

## ADDA

### 训练过程

- 使用源域对映射网络$M_s$和分类器C进行监督训练，优化损失函数$L_{cls}$。

$$\min_{M_s,C} \mathcal{L}_{\text{cls}}(\mathbf{X}_s, Y_s) =$$

$$-\mathbb{E}_{(\mathbf{x}_s,y_s)\sim(\mathbf{X}_s,Y_s)} \sum_{k=1}^{K} \mathbb{1}_{[k=y_s]} \log C(M_s(\mathbf{x}_s))$$

北京大学
PEKING UNIVERSITY

ADDA

## 训练过程

- 固定源域映射网络$M_s$，使用目标域数据和源域对目标域映射网络$M_t$和判别器D进行对抗训练，优化损失函数$L_{\text{adv}D}$和$L_{\text{adv}M}$。

$$\min_D \mathcal{L}_{\text{adv}_D}(\mathbf{X}_s, \mathbf{X}_t, M_s, M_t) = - \mathbb{E}_{\mathbf{x}_s \sim \mathbf{X}_s}[\log D(M_s(\mathbf{x}_s))] - \mathbb{E}_{\mathbf{x}_t \sim \mathbf{X}_t}[\log(1 - D(M_t(\mathbf{x}_t)))]$$

$$\min_{M_t} \mathcal{L}_{\text{adv}_M}(\mathbf{X}_s, \mathbf{X}_t, D) = - \mathbb{E}_{\mathbf{x}_t \sim \mathbf{X}_t}[\log D(M_t(\mathbf{x}_t))].$$

- 使用训练好的目标域映射网络$M_t$和分类器C对目标域数据进行测试

北京大学
PEKING UNIVERSITY

# 实验结果

# CHAPTER FOUR

北京大学
PEKING UNIVERSITY

# 实验结果

北京大学
PEKING UNIVERSITY

# 总结与展望

# CHAPTER FIVE

北京大学
PEKING UNIVERSITY

## 不足与改进方向

1 不足：分类器仅仅使用source domain上的数据进行训练，可能并不能很好地泛化，毕竟仅仅使用source domain训练的分类器可能会带有一定的域特定特征。

2 改进：可以考虑通过权值分享(参数绑定)进行域适应，则是希望将源域和目标域进行共同的映射，映射到所谓的latent feature space，而所做的优化则是优化这种共同的映射，使得latent space更为合理，同时使得源和目标的差异更小，完成域适应。

北京大学
PEKING UNIVERSITY

# 谢谢老师同学！
# 请老师同学们批评指正！

组员：陈亦弘 张鉴心

学院：前沿交叉学科研究院