1. **这篇论文要解决什么问题？要验证一个什么科学假设**

在语义分割上，注释特别昂贵，因此必须标记每个像素，对于在不利因素下像素标记的时间更久，但是UDA可以很好的解决上述问题，但是大多数基于过时的网络体系结构，性能得不到很好的提升。

虽然大多数提出了新的自适应策略，但是都是使用的过时的网络框架。

验证的科学假设：采用新的UDA方法，想要验证新的网络框架在GTA→Cityscapes上方法是否比之前更加好

1. **这篇论文有哪些相关的研究，这些研究是如何分类的，有哪些研究员能够值得关注**

Road segemention

起初是：CNN的引入 Gradient-based learning applied to document recognition.

FCN 的引入 Fully convolutional networks for semantic segmentation.

接下来遵循encoder-decoder design（编码器-解码器的设计）

A deep convolutional encoder-decoder architecture for image segmentation.

U-net: Convolutional networks for biomedical image segmentation.

加入skip和膨胀卷积

U-net: Convolutional networks for biomedical image segmentation.

扩张卷积

Semantic image segmentation with deep convolutional nets and fully connected crfs.

上下文信息和金字塔池化

Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs.

Encoder-decoder with atrous separable convolution for semantic image segmentation.

Grid saliency for context explanations of semantic segmentation.

Pyramid scene parsing network.

注意力模块

Dual attention network for scene segmentation.

Non-local neural networks.

基于transformer，transformer比CNN具有更多的鲁棒性

Understanding robustness of transformers for image classification.

Intriguing Properties of vision Transformers.

ASPP+skip connection 增加了鲁棒性

Encoder-decoder with atrous separable convolution for semantic image segmentation.

skip connections

U-net: Convolutional networks for biomedical image segmentation.

UDA：UDA的方法分为对抗训练和自训练

对抗训练首先调整输入、输出、patch level、鉴别器、类别信息

Domain-adversarial training of neural networks

Generative adversarial nets.

大多数UDA的方法都可以离线预先计算伪标签，训练模型并重复该过程

Model adaptation with synthetic and real data for semantic dense foggy scene understanding

Fda: Fourier domain adaptation for semantic segmentation.

Unsupervised domain adaptation for semantic segmentation via class-balanced self-training.

在线训练伪标签

Prototypical pseudo label denoising and target structure learning for domain adaptive semantic segmentation.

一致性正则化

Fixmatch: Simplifying semi-supervised learning with consistency and conﬁdence.

还有几种方法结合了对抗训练与自训练

Learning texture in variant representation for domain adaptation of semantic segmentation.

Bidirectional learning for domain adaptation of semantic segmenation.

由于数据通常不平衡，并遵循长尾分布导致模型很普通，解决这个问题需要：

1. 重新采样，损失重新加权，转移学习

Re-distributing biased pseudo labels for semi-supervised semantic segmentation: A baseline investigation. (重新采样)

Focal loss for dense object detection. (损失重新加权)

Deep representation learning on long-tailed data: A learnable embedding augmentation perspective.（转移学习）

（2）在UDA中应用重新加权和用于图像分类的类别平衡采样

In stance adaptive self-training for unsupervised domain adaptation.

(3) 知识蒸馏可以充当可以充当新任务的正则化器

Distilling the knowledge in a neural network.

1. **论文中提出的解决方案是什么，关键点在哪里**

论文中提出了DAformer，网络构架由一个Transform的编码器和一个多级上下文感知特征融合的解码器组成。

同时提出三个训练策略，有效训练和避免过拟合到源域上

（1）罕见类别采样（RCS），顾名思义利用模型多去见罕见的类别。原本UDA会对于常见的类别过拟合，那么通过专门去频域上采集一些罕见类别，那么网络就会学习的更加稳定

（2）FD(特征距离的度量)从各种ImageNet上提取知识，以规范源训练，拉近源域与目标域之间的距离。就是在源域仅限于某些类的少样实例，这些实例与目标域有着不同的特征，如果不使用FD，那么，这将导致学习更少的表现力和源域的特征。

（3）在UDA上引入了学习率预热，通过在早期训练中将学习率提高到预期值，可以稳定学习过程，并且将预训练的特征很好的转移到语义分割中。促进了ImageNet预培训的功能转移。

方法的详细介绍：

以往的解决方案：在解决领域差距的问题上，是在源域上使用分类交叉熵损失训练网络，但是问题是会导致目标图像性能低下，因此，网络无法更好的推广到目标域上。目前，有两种解决领域差距的策略。分别是adversial training 和 self-training(ST)。

（1）UDA中的自训练

在本次实验中，我们将采用自训练的方式。

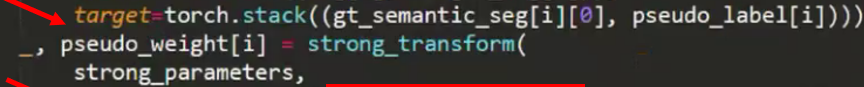
:利用源域训练的一个神经网络

使用源域的图像和有标签的，为了在目标图像上实现良好的性能和未访问的目标标签

第一步：先生成一个伪标签——在自训练中采用了teacher network 的目的是为目标域数据生成伪标签。

这里面的梯度不会反向传播到教师网路中。

借助DACS的代码



第二步：假设向量经过softmax之后，最大概率值大于一个threshold，那么结果有效。因此，在这里设置了一个阈值，其中公式的代表着分子是图像的尺度，分母是每一个图像有效或者无效的情况。

现在变为利用伪标签以及质量估计对目标域上的网络进行训练

借助DACS的代码



第三步：选择在线的ST，因为它只有一个训练阶段，使用在训练过程中基于进行更新。

本实验中采用在线自训练的，在过程中在训练过程中基于进行更新，选用DACS，使用color jtler，高斯模糊，classmix。

关于DAformer的构架

尽管，transformer的自训练和卷积都执行一个加权和，但是计算权重的方式不同。因此，设计了MIT，图像被分为4\*4的，以保证保留更多的信息。

（1）、transformer设计用于产生多级特征图

（2）、我们建议在解码器上利用上下文信息

DAformer不仅使用和只考虑上下文信息，而是使用不同编码器级别的上下文特征。主要是采用1\*1的卷积将嵌入到相同的通道中，并将特诊上采样到，并将他们连起来。

上下文感知的特征融合，使用ASPP来融合所有堆叠的多级特征，并且去掉ASPP中平均金字塔池化

三个训练策略

1. RCS（罕见类别采样）

罕见类采样通过减轻self-training对普通类的确认偏差提高了伪标签的质量。

根据数据采样训练的随机种子，这些类是在训练的不同迭代中学习的。当网络已经学会对普通类产生强烈偏见，这使得很难很少的样本“重新学习”新知识。那么针对于这问题，提出了RCS，更频繁的使用稀有类对图像进行采样，源数据中每个类C的频率可以根据类的像素来计算。

每个类C的频率：

C的样本概率

根据频率越小的类，采样频率越高。温度T的越高导致分布越均匀，T较小导致要关注较小稀有类的。

因此由于一个罕见类（频率小的时候）通常与单一图像中常见类（频率大的时候）更常出现，那么对于罕见类的采样是有意义的（比如摩托车，公共汽车，火车）

1. FD（根据FD的实验结果显示，因为MiT5的编码器性能强大，容易发生过拟合，使用FD进行正则化可以有效的避免性能的下降）

在ImageNet里面还包含来自一些相关的高级语义类的图像，uda难以区分。为了避免ImageNet预训练的有用特征被LS破坏，并且模型对合成源数据的过拟合。先进行正则化，对于UDA模型的bottleneck特征和ImageNet模型的bottleneck特征进行正则化。

ImageNet模型主要训练的事物类包括具有清晰定义的图像，（例如，汽车），因此，只对于包含二进制掩码Mthing描述的事物类的图像区域FD的损失

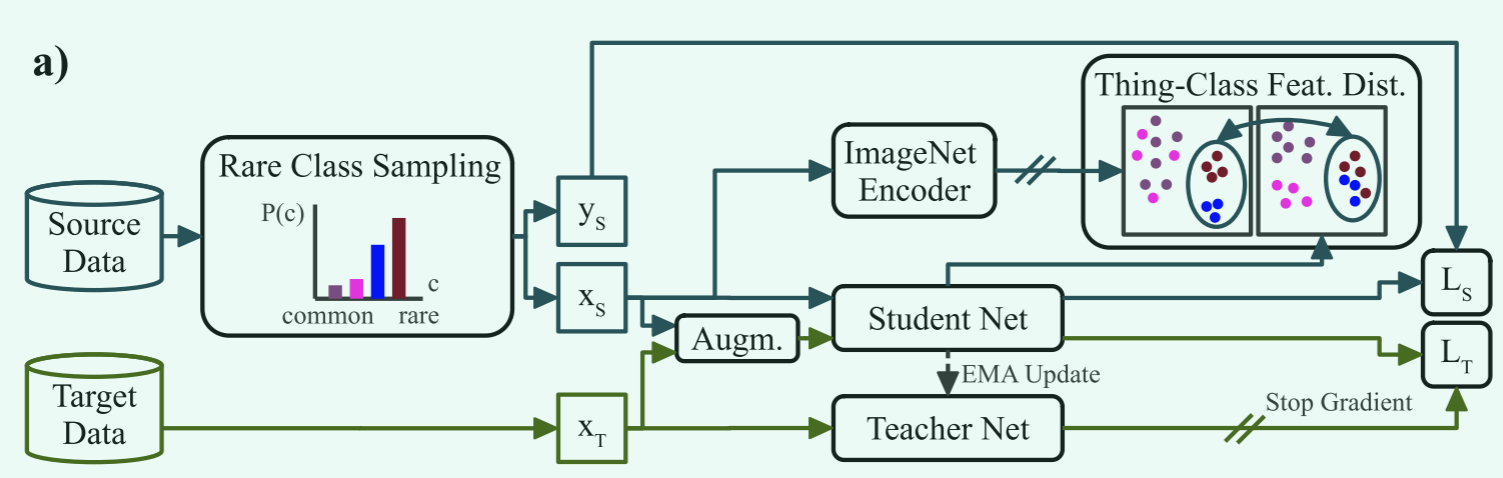
这个mask是从缩小比例的标签中获得的

为了将标签采样到瓶颈特征大小，对每个通道应用patch size的平均池化，当一个类超过比率r时，则该类被保留。由于稀有类通常与单个图像中多个公共类共同出现

这确保了只要包含主导物类的瓶颈特征像素

（3）学习率warmup

为了提高训练的稳定性，在最初几轮迭代时，采用较小的学习率，等梯度下降到一定程度之后，在恢复学习率，这种方法称为学习率预热



Source data：源域，在源域的数据集上，进行RCS，之后得到（XS，YS）就是在源域的数据和标签。

Target data：目标域，在目标域上只有数据XT没有标签。

关于数据方面，在（源域）XS和（目标域）XT上数据进行通过Augm进行数据增强。

之后XT经过Teacher Net得到伪标签。

Student Net：相当于，Student Net用来做网络的更新

teacher Net：用来生成伪标签

红色框标注的地方采用的是mixing的方式。根据DACS里面的方法，Ym就是

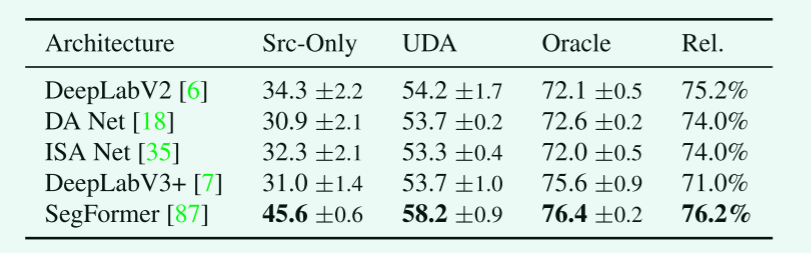
Xm就是，得到（Xm, Ym）。

关于Ls和Lt的理解，个人认为是：

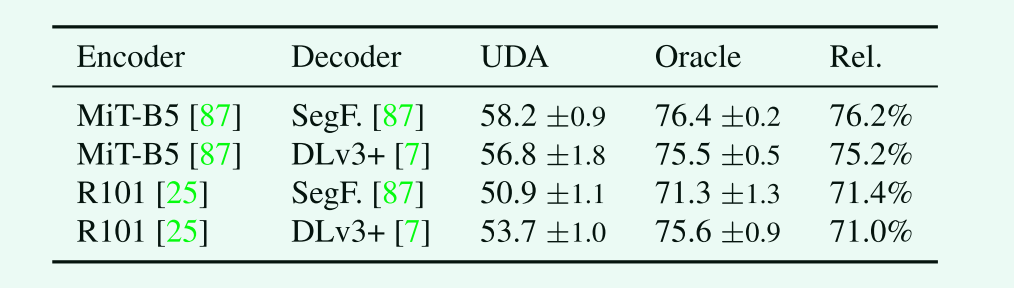
首先，通过student net得到源域的预测结果，同时通过student net得到源域的预测结果，在源域的预测结果和源域之间有一个损失Ls。之后通过teacher net得到混合之后的预测结果，在混合域的预测结果和混合域之间有一个损失LT。在ImageNet的编码器里面用到了FD。

1. 论文中的实验室是如何设计的？各个实验分别得到了什么结论

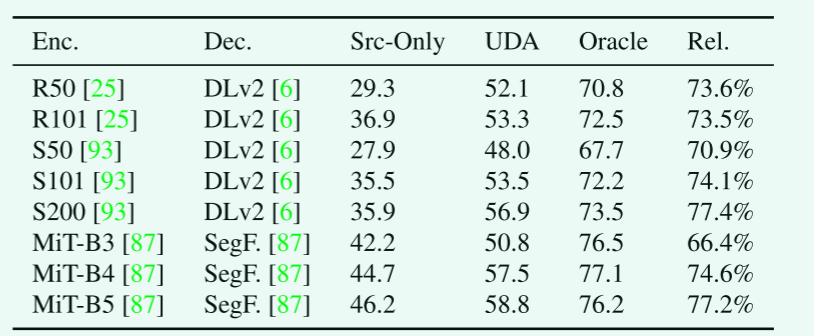
首先与之前几个常见的分割框架UDA的性能进行比较

****

UDA上大部分工作使用DeepLabv2和restnet101 backbone。可以从表中看出即使很高的Oracle，也没有提高UDA的性能，观察DeepLabV2和DeepLabV3+可以看到，但是segFormer不管是Oracle还是UDA上都提高了大部分的性能。那么segFormer更适合UDA



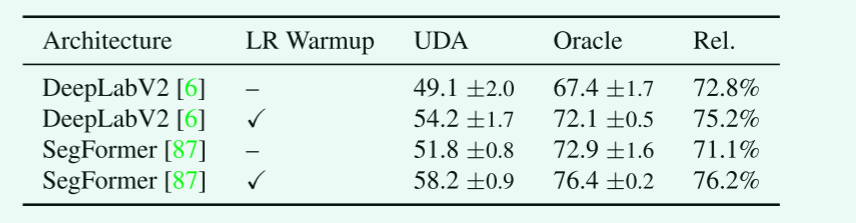
想要知道SegFormer中哪一个部分起到了关键的作用。将编码器和解码器与ReaNet101和DeepLabv3交换。由于SegFormer的MiT编码器输出步长为32，而DeepLabv3解码器的输出步幅为8，当与DeepLabv3和解码器结合使用的时候，我们对于SegFormer的bottelneck的特征，进行\*4的采样。因此对编码器和解码器都进行了替换实验。表二中看MiT-B5的编码器的作用更大。

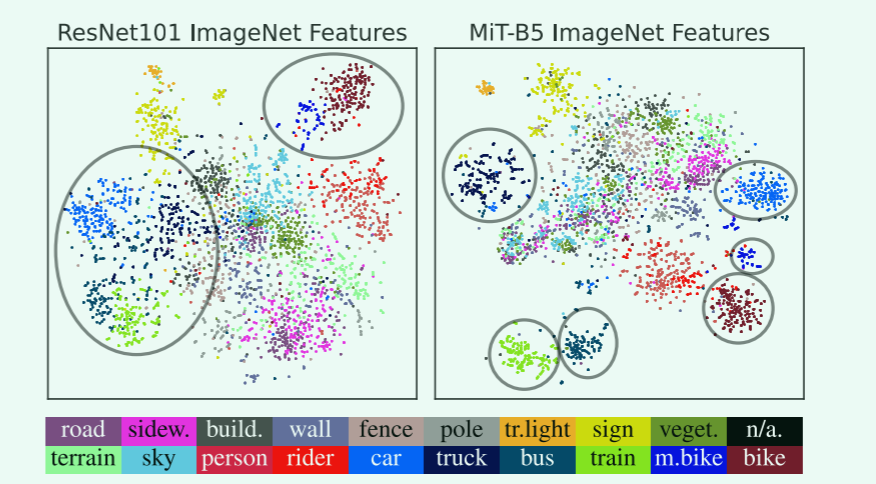


选择了不同的编码器和解码器进行实验，进一步了解编码器对UDA性能的影响。从表中可以看出更深层次的模型可以更好的首先Src-only，所以采用更深的模型是由意义的。

消融实验

（warmup）学习率的作用，与没有加学习率的效果呈现的提高，说明学习率的重要性

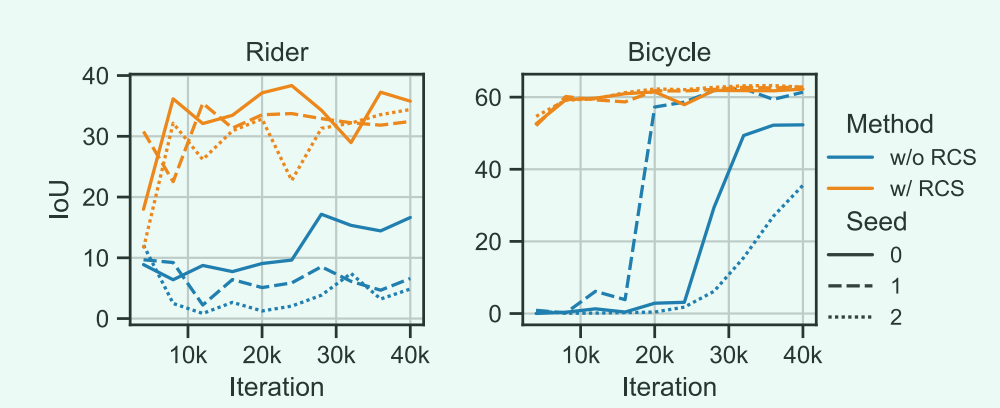




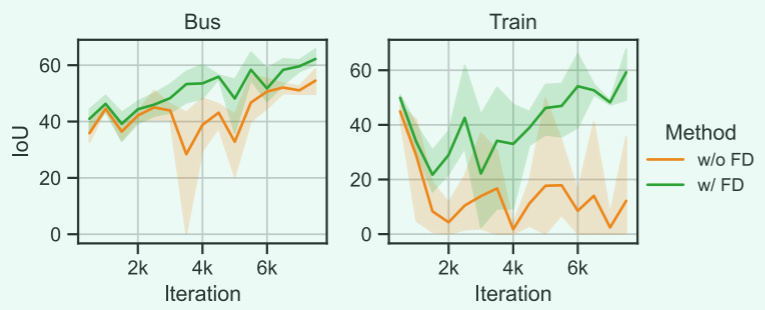
这个可视化了目标域上的ImageNet的特征，即使ResNet的结构的物体类稍微好一点，但是ResNet101在特征较为相似的类上还是很难区分，比如m.bike和bike上。但是在MiT-B5上就有较为明显的区分，比如car，bike就能很清晰的被区分出来。

（RCS）罕见类别采样(当UDA训练SegaFormer)

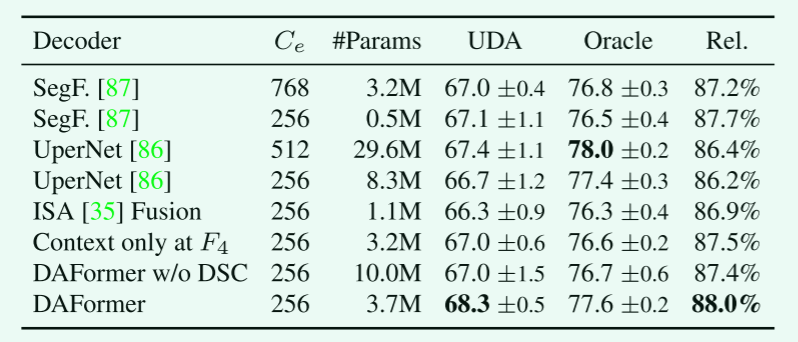
关于罕见类别采样的实验-某些类的性能取决于数据采样的随机种子，图中蓝色IOU曲线，在自行车类的IOU在不同种子的不同迭代中开始增加。假设这是由采样顺序引起的，特别是对相关rare class进行采样。此外IOU提升越晚，该类最终的IOU最差。这可能是早期迭代中积累的自我训练的确认偏差。结论，对于UDA来说尽早学习稀有类别很重要。橙色的线表示RCS会导致rider和bicycle的IOU提前增加，最中IOU独立于数据采样与随机种子无关，这证明了我们之前的假设，一个罕见类的采样对于正常学习这些类很重要。



FD-虽然RCS提高了性能，但事物类的性能可以进一步提高，因为在UDA训练后，一些在ImageNet特性中分离的相当好的对象类混合在一起。在调查早期训练的IOU时，观察到train class的早期性能下降。假设强大的MiT编码器过度适用于合成域，当使用FD进行正则化时，避免了性能下降。



Daformer的解码器效果



可以看出在解码器上DAformer有明显的提升。

5、用于定量评估的数据集是什么？代码开源给出链接

目标域使用Cityscapes street scene dataset 包含2975张训练图像和500张验证图像，分辨率为2048\*1024.

源域使用GAT数据集，包括了24966张分辨率为1914\*1052的合成图像

或者使用Synthia dataset数据集，包括9400张合成图像，分辨率为1280\*760

<https://github.com/lhoyer/DAFormer>

**6、这篇论文到底有什么贡献？（三句话说明新的问题）新的地方**

这篇论文提出了新的基于UDA的框架DAformer，由transformer+多级上下文感知特征融合编码器组成，并结合三个训练策略（RCS+FD+学习率预热）

**7、下一步基于它做什么？有什么工作可以深入**

可以基于UDA新提出的DAformer模型去解决关于恶劣条件下的图像分割问题

（1）DAformer的解码器由作者设计，我们可以更改解码器，看是否有效果，当通过1\*1卷积将每个嵌入到相同数量的通道中，可以采用转置矩阵进行上采样之后在拼接。

（2）更换数据集，Daformer是否能够很好的运用在恶劣场景下

（3）可以改进数据增强的方式