INTRODUCTION

visio是微软公司的一款框图制作软件，通过拖选visio预置的图形进行组合拼接，制作所需要的框图。Visio中数以万计的丰富图形素材库给予了用户很大的便利与选择空间，但也带来了些许问题。许多用户使用visio时难以找到合适的图标，因为他们只知道自己想要什么类型的图标，而不知道这些图标在visio里的具体分类与名称。相比尝试词条来检索图标，如果能直接通过绘制出大致想要的图标款式，而软件直接列出匹配得可能的选项，就可以减少许多的麻烦， 即便是对于常用的基础图形，直接将它画出来也相比去图形栏中寻找更加方便，这也是what initially motivates us。

更进一步，不止拘泥于单个基础图形的识别，若能直接通过手绘草图直接进行建模与参数化过程，直接得到完整的最终框图，将会极大地提高相关从业人员的工作效率，设计者将省去挑选图形、组合、设置尺寸、排版等一系列繁琐的工作，只需要对模型的结果进行微调。这将使设计者能够更好利用自己作图的潜能与自由，在利用参数模型中提供的检索和生成方案的同时，增加绘制草图的灵活性。

如何利用计人工智能技术实现自动且高效地对手绘草图进行精准地语义解析与识别，进而提升相关从业人员的工作效率，是一项极具意义和挑战性的前沿课题。 为了应对这一挑战，我们制作了一个Visio图案的手绘草图数据集，引入了Visio-Sketch-Network，一种交互式用于早期流程框图设计的、基于手绘草图的生成系统。用户不需要操作繁琐的编程式绘图页面，相反，我们的模型能自动识别，使未经训练过的用户就能快速绘制出复杂的Visio图。

RELATED WORK  
  
在计算机视觉领域，手绘识别是一个较为基础但又极其重要的问题。随着触屏等交互设备越来越普遍，人们能够更方便的使用直观的sketching来传递信息，因而，对于手绘图形的识别的实现需求越来越大。有限的可用数据集，不同人绘制同一个物品的差异性，以及手绘图片的稀疏性，让手绘体识别依然是一个具有挑战性的工作。

在应用领域，手机、平板电脑等电子设备上，一些厂商已经实现了类似的功能。例如，在ios系统的手写输入法中，通过手绘一些图案，预测输出栏会给出相匹配的图标。虽然这些识别算法对于文字输入的识别效果极佳，但对于sketching图案的识别常常无法给出令人满意的结果。

在3D设计建模领域，Mohammad Keshavarzi提出了基于Autodesk与卷积神经网络的Sketch生成方案SketchOpt，一定程度上化简了设计前期用户基于性能的生成设计系统作图的操作量，能通过草图直接生成对应设计方案。  
在算法方面，更多的工作关注与较复杂手绘图形的草图分割与语义理解，对于草图的处理主要有sequence-based[sketch-r2cnn,2018]和image-based、以及pointset-based[sketchgcn,2020]的编码方法。在此之外，也有更多基于笔画节点信息的编码方式,例如2020年的multi-graph transformer，将笔画节点的坐标信息与位置、状态信息升维、拼接作为模型输入，在quickdraw数据集上达到了sota表现。不过这些工作都基于较为复杂的，包含更多特征的复杂手绘classes，而对于较为简略，几何形状单一的手绘图型的分类工作，这些道路需要进一步检验  
  
数据集是另一个值得关注的工作。草图数据集不同于自然图像数据，人们只需要通过拍照便可轻松获得传统RGB图像数据集，而手绘草图则需要一笔一划创作而得，时间成本和收集难度远大于传统图像。

关于常用标签、事物的数据集有谷歌的quickdraw dataset，tu-berlin数据集，以及imagenet的手绘图数据集等。但是，在visio中，我们最初需要的，关于基础图型的数据集却少之又少。事实上，基础图型中有许多ambiguous classes，在识别时造成混淆，例如椭圆和圆，在未经过绘画训练的手绘者的误差影响下，很难判断绘图者的真正意图。

我们的contribution可以总结如下：

1. 制作了一个基础图形数据集，对基础图形手绘数据进行了补充
2. 提出了一种基于不同尺寸图像的草图语义解析方法，在我们的数据集上准确率优于（XXXX，需要有个准确实验）
3. 结合用户对visio软件的使用需求，运用机器学习卷积神经网络的方法制作了一款插件

METHODOLOGY

1、Dataset

利用微信、二维码利于传播、应用广泛的特点，我们设计了专门的收集数据微信小程序，通过数据云托管平台，收集了348位学生、教师、公务员、企业家等来自不同领域、不同年龄段人群的手绘，制作了一个关于基础图型的手绘数据集，共有30类，包含如圆、椭圆、十字形、五角星、矩形等常用的基础图型，每一类约有80张，共计2503张sketch，这符合我们对于最初设想的功能的基本需求，

与基础图形库不同，在Visio中，特定领域的图形素材库间往往区别极大，例如电路库中，有电容、电感、电磁继电器等一系列专业的符号表示，计算机科学领域也有服务器、路由器等不同的设备。在此基础上，我们又针对性地设计了额外20类，包括计算机、文件夹等需要一些domain knowledge的类别，并收集了来自20位计算机领域学生的，各100张，合计2000张的手绘草图。这与Visio图形素材细粒度地划分类别、领域相符合，也与如今小样本机器学习数据集越来越针对化的趋势相吻合。

最后，我们又对各个类别的图片进行了筛选和补充，共计50大类5303张图片。

2、CNN图像识别

之后，我们提出了基于不同尺寸图像的草图语义解析方法，即将输入的图像resize为多种尺寸并输入不同CNN，最后通过Bayesian method综合输出得到结果的方案。训练过程中，我们使用了随机反转、旋转等数据增强手段，并对sketch-A-Net经过fine-tuned，我们的网络在自己的数据集上可以达到百分之84.6的准确率，在TU-Berlin上能达到…Googlenet上能达到…（自己电脑训练不起来）

3、图像分割——综合框图生成

CASE STUDIES

预设实验1、不同网络在该数据集下的识别正确率表现

2、用该数据集的预训练模型，正确率较random和其他数据集高。

3、完整草图生成展示

DISCUSSION  
在我们的研究中，我们已经完成了一个基础图形识别算法到UI的完整系统搭建。相比于其他工作，我们对于相似草图类别区分的性能更加看重，因为这对于Visio制图匹配的任务十分重要。如前文所言，在已经完成的基础图形识别功能中，我们已经面对了此类问题，例如平行四边形与菱形，圆与椭圆的区分。对于笔画单一的基础图形，我们使用了多尺度特征图输入的方案，只利用了图像信息，在这个任务上达成了不错的效果。  
  
  继续推进本项目，我们还需要面对的挑战与需求主要有如下几条：  
     首先，在我们最终想实现的构想中，一个更完善的匹配系统会面临类别相近的问题。在visio中，每一个类别都可以有许多细分小类，例如，对于一个厢式货车作为基础的图标，visio里有对应的警车，救护车，快递车等，用户知道这些分类的存在，但在使用绘图表示时，希望他们能够在手绘时精确体现细节的特征是不现实的，如何匹配这些细分类会是一个较为困难的问题  
     同时，即便在这些细节被用户较为得体地手绘出来，针对这样的细粒度图像分类问题，我们可能需要引入一些在手绘图上挖掘Discriminative region的机制，来实现较好的效果。  
     其次，对于想要实现的功能，最初由我们定义的类别是不够的，用户往往会使用到更多未训练过的类别。我们可以与用户协议，以便在用户绘制出已有类别之外的手绘时，将其搜集为训练数据，然而，与我们专门搜集的类别相比，一个之前未定义的类别可能只能搜集到少量的数据。如果想要我们的功能可以适时的更新类别，以满足更多用户需求，就需要使用小样本学习来完成这个目标。  
  
     作为一个功能性软件，非常重要的一点是具有足够的鲁棒性。具体而言，对于存在ambiguity，多重理解，或者绘制不佳的手绘，我们希望尽量的得出正确的结果，这对于我们模型的泛化能力提出了要求。与针对真实图像的CV识别算法的鲁棒性评价不同，我们需要考虑手绘图情形下噪声和扰动是以什么样的形式输入的。因而，除了我们已经使用过的随机翻转，我们可以引入更多的数据增强手段，例如将手绘图的部分区域遮挡，部分笔画擦除、或者比例变形等，将这些干扰加入训练和测试可以进一步提升算法的性能。鲁棒性的另一方面干扰在于绘图理解的不同，当我们面对比基础图形更复杂的绘图时，对于同一标签，存在完全不同的绘图方式，对于这类问题，我们需要测试算法的实际表现。  
最后，对于我们扩充新类别的需求，我们也要考虑添加相似的类并进行迁移学习后对于原有类的影响，在实现时非常有必要对结果进行分析测试。

CONCLUSION

在本研究中，我们提出了一种基于多尺度cnn的基础手绘图形识别算法，并提出了一个包含50个标签基础图型数据集。该数据集可以被应用于visio等制图软件的用户辅助功能，使得用户可以直接通过手绘来获取匹配的基础图型，由匹配单一基础图型更进一步，对于用户绘制出的一张由多种基础图型组成的构想输入，识别其中的图形类别、位置、几何关系参数，直接得出组合好的设计框图，可以使得制图效率大幅提升，尤其是那些对于制图工具软件不熟悉的人。我们也将进一步从各方面提高算法的准确性与鲁棒性。最后，我们的研究现阶段还有部分限制，第一，对于一些基础图型类别，用户手绘导致的ambiguity还无法消除;其次，目前对于单个基础图型的识别，在实用性上有些欠缺，这些问题会在后续工作中得到解决；最后，Visio拥有成千上万种类别，且不同种类间有的极其相似，想要实现大量类别的识别，需要相当大的数据量，如何实现sketch的小样本学习并保持高正确率，这将是未来极大的挑战。

参考文献