

## 集群使用 V1.0

### 1、 文件管理（该文件多为 windows 环境所写，掌握思想，触类旁通）

每个人在使用电脑的时候都要在某个盘（可能有的盘满了，需要去另一个盘）里创建自己的个人文件夹，除非特殊情况，必须将使用的数据或软件等放在以自己名字命名的文件夹下，并且一定是英文，比如 wuwei。（因为我们大多是图片处理，而 cv2，对输出路径有要求不能是中文，并且有的计算机软件，python 库等等都不识别中文，可以避免暗坑）。

#### 一、自己独立文件夹命名方式

在/home/sjwlab 创建自己独立文件夹：

如 董秋彤，命名为 dongqt （姓的拼音字母全称+名的第一个拼音字母）

```
dongqt liny
```

#### 二、文件命名方式

所有的文件夹除了 dataset、project、output 为强制使用，其他文件夹根据情

况使用，并可以根据情况创建新的文件夹

在 wuwei 下创建文件夹（后有详细解释）：

dataset（存放原始数据），

按数据集创建文件夹

project（项目代码），

按项目创建文件夹

preprocess（预处理数据），

按项目创建文件夹

output（输出结果），

按项目创建文件夹

software（存放软件），

可以根据情况对结构进行修改，但要确保，删除的时候不会出错，需要

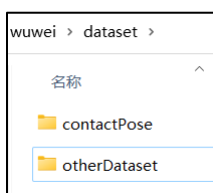
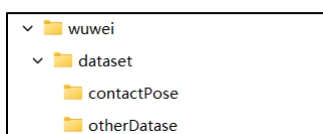
找到某个文件的时候可以快速找到

使用完数据以后，确定没用以后，使用绝对删除（**shift+delete**），要

确认删除的文件，不要删错，如果有预处理的数据也要删除。

## 1、存放数据 (dataset)

假设要存放的 contactPose 的数据。那么就要在 wuwei 下创建一个文件夹 dataset，并在 dataset 里创建 contactPose 文件夹，并放入数据，并且**一定要创建一个 readme.docx 文件**来说明这个数据来自哪里是做什么的，包括一些特殊的说明等等。例子如下图所示：

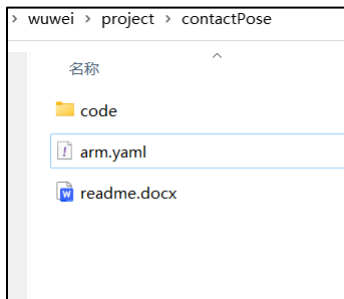
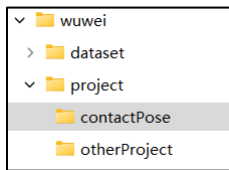


readme.docx 内容可以类似如下



## 2、存放代码（project）

假设代码项目的名称是 `contactPose`，则需要在 `wuwei/project` 下创建一个 `contactPose` 文件夹。并且同样可以写一个小的 `readme.docx`，里面写使用的 `conda` 环境名是哪个，如果使用命令行启动还可以写上启动命令，配置文件地址，配置文件参数含义，要注意什么等等



Readme.docx 的内容可以如下

环境: contactPose

启动使用了命令行启动, 启动命令为

mlagents --config=D:\wuwei\project\contactPose\arm.yaml --gpus=4

配置文件是: "D:\wuwei\project\contactPose\arm.yaml"

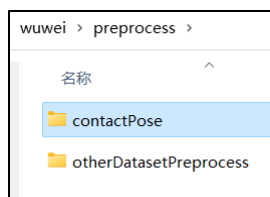
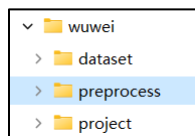
这里还可以写一下配置文件里的参数含义等等比如

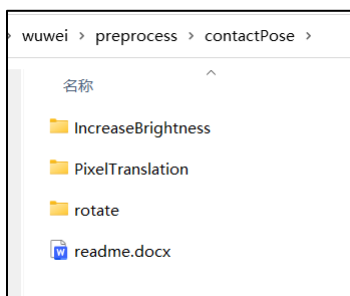
```
E: > UnityProjects > gym1 > Assets > config > ppo > ! arm.yaml
1  behaviors:
2      arm:
3          trainer_type: ppo
4          hyperparameters:
5              batch_size: 64
6              buffer_size: 12000
7              learning_rate: 0.0003
8              beta: 0.001
9              epsilon: 0.2
10             lambd: 0.99
11             num_epoch: 3
12             learning_rate_schedule: linear
13         network_settings:
14             normalize: true
15             hidden_units: 128
16             num_layers: 2
17             vis_encode_type: simple
18         reward_signals:
19             extrinsic:
20                 gamma: 0.99
21                 strength: 1.0
22         keep_checkpoints: 5
23         max_steps: 500000
24         time_horizon: 1000
25         summary_freq: 100
26
```

Learning\_rate\_schedule 是学习率的变化形式, linear 代表的是线性衰减直到结束的时候学习率会变成 0

### 3、预处理数据 (preprocess)

有的数据需要进行预处理, 而不是在训练的时候一边处理一边训练 (这可以大大减少训练速度) 比如, 数据增强, 预先增强完就可以放在这个项目下。假设要对 contactPose 数据集进行数据预处理, 那么就需要在 wuwei 下创建 preprocess 文件夹, 并在 preprocess 里创建 contactPose 文件夹, 不同的预处理还要在 contactPose 里创建多个文件夹。如下所示





Readme.docx 文件内容可以类似如下

数据地址  
D:\wuwei\dataset\contactPose

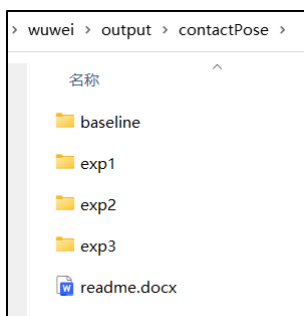
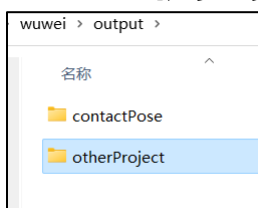
已经数据增强了，  
旋转（D:\wuwei\preprocess\contactPose\rotate）  
像素平移（D:\wuwei\preprocess\contactPose\PixelTranslation）  
增加亮度（D:\wuwei\preprocess\contactPose\IncreaseBrightness）

增强操作	旋转	像素平移	增加亮度
增强程度	旋转 20°	向右平移 3 个像素	增加十点亮度
增强结果	lou 从 0.6 到 0.8	lou 变化不大从 0.6 到 0.61	lou 从 0.6 到 0.9

根据情况添加别的备注

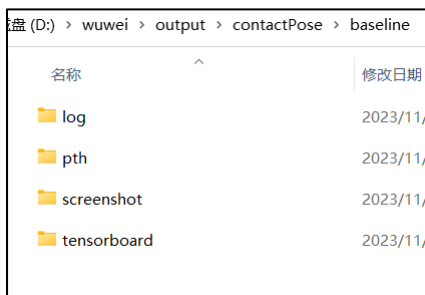
#### 4、输出 (output)

假设 contactPose 有了输出。那么就需要在 wuwei/output 下创建 contactPose，并在 contactPose 里创建多组实验的文件夹，如下

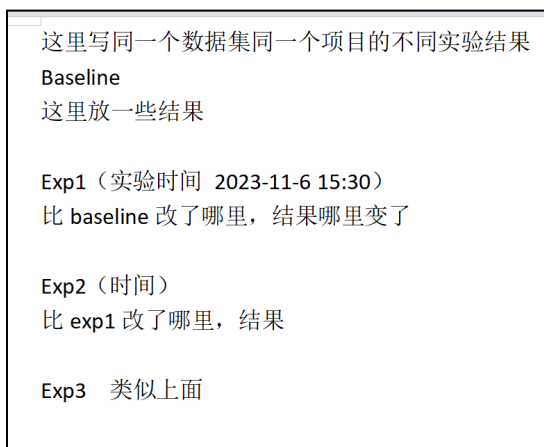


而每次实验都要注意保存相关信息，比如 pth 参数（一定要保存），截图，tensorboard

**文件 (这个一定要保存)**, 可以在 log 里存一下其他信息



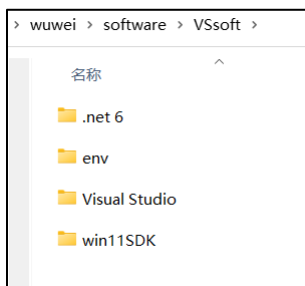
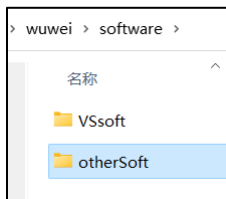
其中 readme.docx 要写每次实验的情况已经与上次实验(或者是对基线实验的改动与结果), 可以类似如下



## 5、软件 (software)

假设要安装 Visual Studio, 但是会绑定很多其他附属软件

那么可以在 wuwei/software 里创建 VSsoft, 如下



## 2、 算力负载登记文件在线文档，

是记录的所有的设备的连接方式，包括有 todesk 可以链接到有视图的页面，也可以使用 ssh 直接使用。推荐使用 ssh 链接，稳定快速，todesk 可以使用 pycharm，最好的形式是取少量数据来让模型能够顺利运行，再将代码与数据拿到服务器上运行。大量数据传递可以使用 bypy。

## 3、 Linux 简单使用 (ssh)

链接：windows 可以使用 xshell 或 mobaxterm

mac 可以直接使用自带的终端命令为：ssh -p 端口号 用户名@主机（也就是 host 域名） 然后输入密码即可

1、ls 显示当前目录下的所有文件与文件夹

```
[(base) clks@clks-X12DPi-N-T-6:~/sjwlab$ ls  
dongqt liny
```

2、cd 切换路径 后跟路径名 可以使用 tab 补全路径（可以补全文件名字也可以是文件夹名）

```
[(base) clks@clks-X12DPi-N-T-6:~$ cd sjwlab/  
[(base) clks@clks-X12DPi-N-T-6:~/sjwlab$ ls
```

3、pwd 可以显示当前的路径

```
[(base) clks@clks-X12DPi-N-T-6:~/sjwlab$ pwd  
/home/clks/sjwlab
```

4、cat 查看文件内容

```
[(base) clks@clks-X12DPi-N-T-6:~/Desktop/iGuge$ cat README.md  
# iGuge  
iGG谷歌访问助手  
离线下载器  
URL: https://iguge.app/
```

其他 Linux 命令询问 chat 都可以得到

conda 已经安装

conda 和 pip 的使用与 Windows 的命令行一样



4、bypy 的使用，前面里面有使用，更换账号，搭配多线程下载器来加速下载，最后也有安装的教程。

## Linux 下使用 bypy 下载百度网盘数据

有使用与安装，安装放在后面，没安装的可以参考，已经安装的可以直接使用

使用：








使用

bypy list

可以查看所有“我的网盘/我的应用数据/bypy/”下的数据如 (D 代表文件夹, F 代表文件)：

```
(base) aa@aa-iGame-Z790D5-FLOW:~/zhangcx$ bypy list
/apps/bypy ($t $f $s $m $d):
D bypyTest 0 2024-03-28, 11:32:34
D lstm-0322 0 2024-03-22, 23:05:36
D yyl 0 2024-03-27, 15:13:04
F 312121_2048-512_w30_lr001-1-ddp.py 4551 2024-03-23, 01:07:23 0020006c8nf1192eaf765439182a4e5a
F 312121_2048-512_w30_lr001-3-ddp.py 4552 2024-03-24, 00:44:20 0f4fd1d73n1075e86a138be24ba5e81b
F model_checkpoint_epoch_10.pth 89209952 2024-03-24, 00:27:04 7cc0da081vcdcecee1d3c1b2d8d63dd8
F model_checkpoint_epoch_20.pth 89209952 2024-03-24, 00:17:08 09d19490fh0a8e8bcf058b306dba5cfff
```

这与百度网盘里是对应的

<div><div></div><div></div><div></div><div></div></div> 我的网盘 > 我的应用数据 > bypy >		
<input type="checkbox"/>	文件名	修改
<input type="checkbox"/>	 yyl	20%
<input type="checkbox"/>	 lstm-0322	20%
<input type="checkbox"/>	 bypyTest	20%
<input type="checkbox"/>	 model_checkpoint_epoch_20.pth	20%
<input type="checkbox"/>	 model_checkpoint_epoch_10.pth	20%
<input type="checkbox"/>	 312121_2048-512_w30_lr001-3-ddp.py	20%
<input type="checkbox"/>	 312121_2048-512_w30_lr001-1-ddp.py	20%

最终举例如：

```
bypy downdir bypyTest --downloader aria2
```

会在当前目录下下载 “我的网盘/我的应用数据/bypy/bypyTest” 这个文件夹所有的文件 (注意在下载时候只需要打 bypyTest 即可，bypy 会自动补上 “我的网盘/我的应用数据/bypy/” )

安装：

bypy 是基于 python 的依赖

需要安装 requests 和 bypy

```
pip install requests
```

```
pip install bypy
```

安装完以后使用

```
bypy info
```

会触发输入 token 样式如下

Please visit:

```
https://openapi.baidu.com/o1uth/2.0/a2thorize?  
client_id=q8WE4EpCsau1oS0MplgMKNBn&response_type=code&redirect_uri=oob&scope=basic+netdisk
```

And authorize this app

Paste the Authorization Code here within 10 minutes.

Press [Enter] when you are done

将链接 (图中红框) 复制，并且打开网页在地址栏输入这个链接，会提示登录百度网盘账号 (如果没有登录的话)，会获得一个 token，输入 token 以后会出现类似如下的样式：

```
028b67251f30e649a29f9dd0acf6fbd4
```

```
Authorizing, please be patient, it may take upto 300 seconds...
```

```
Authorizing with the Heroku server ...
```

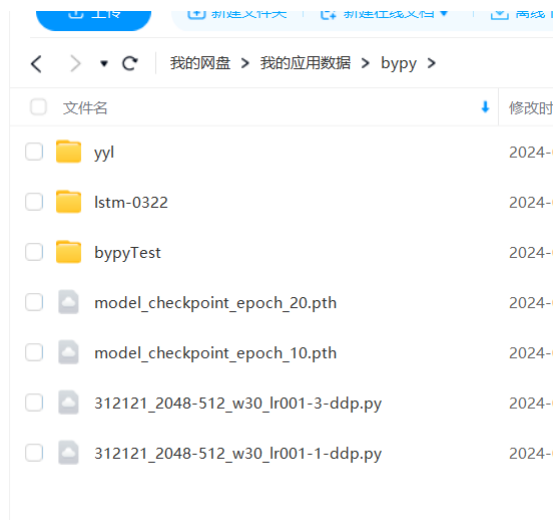
```
Heroku server failed, authorizing with the Heroku1 server ...
```

```
Successfully authorized
```

```
Quota: 5.007TB
```

```
Used: 1.215TB
```

表示配置成功，之后会 bypy 自动在百度网盘里创建文件夹如下



可以将需要的数据放在该目录下，**一定要**创建自己的名字的文件夹  
另外 bypy 的本身下载速度不高，这里使用 aria2，安装命令如下：

```
sudo apt-get install aria2
```

后面可以使用 aria2 来下载数据理论上 50s，5G，[看完整举例使用](#)

更换账号：

需要删除/home/.bypy 里的 bypy.json，并再次在终端里输入 bypy info 输入新的 token 即可

