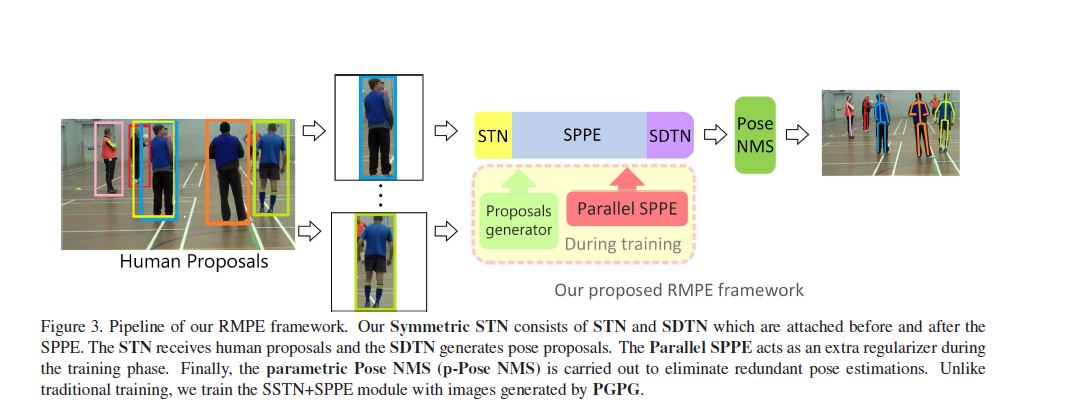
**AlphaPose人体姿态估计总结**

（RMPE: Regional Multi-Person Pose Estimation）

它也是一种自上而下的人体姿态估计方法，提出了一种专门用于区域多人姿态估计的框架（RMPE），主要由三部分组成：

1. SSTN（对称空间变换网络）: 在不精准的区域框中提取到高质量的人体区域
2. PNMS（参数化姿态非极大抑制）：去除冗余姿态
3. PGPG（*姿态引导区域框生成器*）：增强训练数据

整体框架图如下所示：



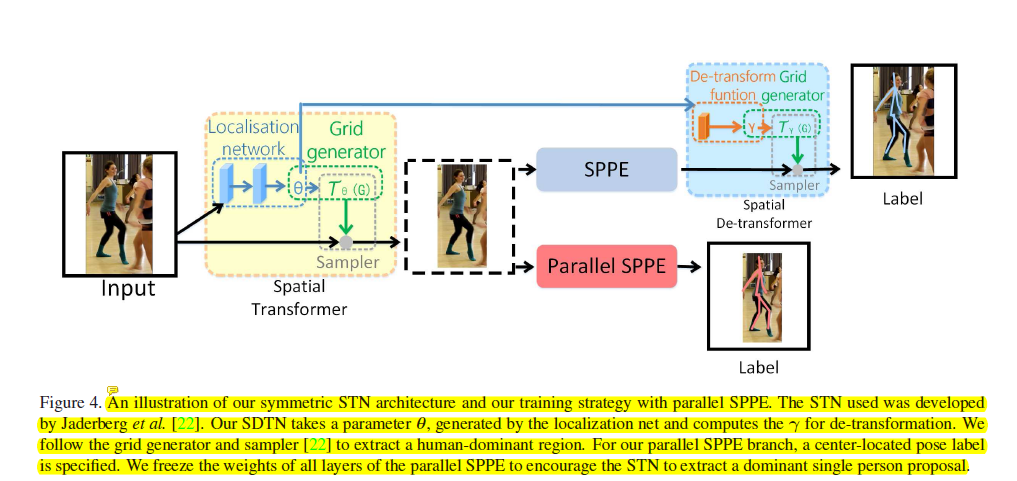
**SSTN=STN+SPPE+SDTN**

大概流程：（1）先通过目标检测算法，将一幅图像**S**中的人分别框出来，此时相当于得到单个人体的原始图像**A**；（2）将单个人体的原始图像输入STN（空间变换网络），通过一些翻转平移等变换，可得到一个高质量的人体框**B**（目标图），即此时的人体框更准确；（3）将此时高质量的人体框图**B**输入SPPE（单人姿态估计），可得到单人姿态估计线条图**C**；（4）将**C**图输入SDTN（空间反变换网络），即相当于将姿态估计线条映射回原始图像**A**，因为可能有冗余姿态，所以加上PNMS，这样最后可在最原始图像**S**中表现出每一个人体的姿态估计线条。

**重点：**

**Symmetric STN and Parallel SPPE**（**对称空间变换网络，并行单人姿态估计**）

目标检测算法得到的人体区域框不是非常适合SPPE，因为SPPE算法是专门针对单个人的图像进行训练的，并且对于定位错误十分敏感。通过微小变换、修剪的方法可以有效的提高SPPE的效果。SSTN+Parallel SPPE可以在不完美的人体区域检测结果下有效的增强SPPE的效果，结构如图所示。



**图解：**表示了SSTN + Parallel SPPE模块的结构，SDTN结构接收一个由定位网络生成的参数θ，然后为反向转换计算参数γ。我们使用网格生成器和采样器去提取一个人的所在区域，在Parallel SPPE中，制定一个中心定位姿态标签。我们固定Parallel SPPE的所有层的所有权重来增强STN去提取一个单人姿态区域。

**STN and SDTN**（spatial de-transformer network，空间反变换网络）。STN能很好地自动选取ROI，使用STN去提取一个高质量的人体区域框。（ROI是Region of Interest的简写，指的是在“特征图上的框”； 1）在Fast RCNN中， RoI是指Selective Search完成后得到的“候选框”在特征图上的映； 2）在Faster RCNN中，候选框是经过RPN产生的，然后再把各个“候选框”映射到特征图上，得到RoIs

STN（[论文阅读](https://blog.csdn.net/m0_37644085/article/details/82632517)）可以被安装在任意CNN的任意一层中——将原来的一层结果U，变换到了V，中间并没有卷积的操作。通过U到V的变换，相当于又生成了一个新数据，而这个数据变换不是定死的而是学习来的，即然是学习来的，那它就有让loss变小的作用，也就是说，通过对输入数据进行简单的空间变换，使得特征变得更容易分类（**往loss更小的方向变化**）。另外一方面，有了STN，网络就可以动态地做到旋转不变性，平移不变性等原本认为是Pooling层做的事情，同时可以选择图像中最终要的区域（有利于分类）并把它变换到一个最理想的姿态（比如把字放正）。在得到高质量的人体检测框后，可以使用现成的SPPE算法来继续高精度的人体姿态检测，在训练阶段中，SSTN和SPPE一起进行fine-tuned。

**注意：**不准确的检测框经过STN+SPPE+SDTN，STN对人体区域框中的姿态进行形态调整，输入SPPE做姿态估计后得到姿态线（人体骨骼框架），再用SDTN把姿态线映射到原始的人体区域框中，以此来调整原本的框，使框变成精准的。

**Parallel SPPE。**为了进一步帮助STN去提取更好的人体区域位置，在训练阶段添加了一个Parallel SPPE分支。这个Paralell SPPE和原来的SPPE用同一个STN模块，和SPPE并行处理时候，忽略SDTN模块。这个分支的人体姿态标签被指定为中心,更准确的说，**SPPE网络的输出直接和人体真实姿态的标签进行对比**。在训练过程中会关闭Parallel SPPE的所有层（？？？？），我们固定这个分支的权重，**其目的是将中心位置的位姿误差反向传播到STN模块**。如果STN提取的姿态不是中心位置，那么Parallel SPPE会返回一个较大的误差。通过这种方式，我们可以帮助STN聚焦在正确的中心位置并提取出高质量的区域位置。Parallel SPPE只有在训练阶段才会产生作用。

**Discussions。**Parallel SPPE可以看作是训练阶段的正则化过程，**有助于避免局部最优的情况**（STN不能把姿态转换到提取到人体区域框的居中位置）。但是SDTN的反向修正可以减少网络的错误进而降低陷入局部最优的可能性。这些错误对于训练STN是很有影响的。通过Parallel SPPE，可以提高STN将人体姿态移动到检测框中间的能力。

**自己训练和测试：**

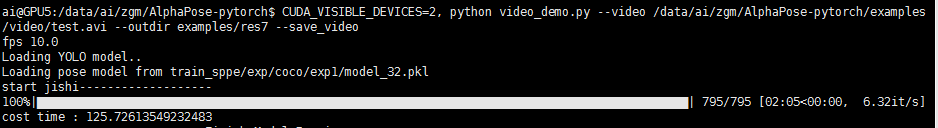
进行测试时，需要两个文件，分别是目标检测的模型（暂时用的是yolo3）和SPPE的模型，目前SPPE部分正在训练过程中,使用了两个gpu，目前训练到了33个epoch，每一个epoch之后都保存的是.pkl文件。

训练过程如下：



**对视频进行测试**：

利用训练好的model\_32.pkl以及目标检测模型yolov3-spp.weights对10fps的原始视频进行测试：

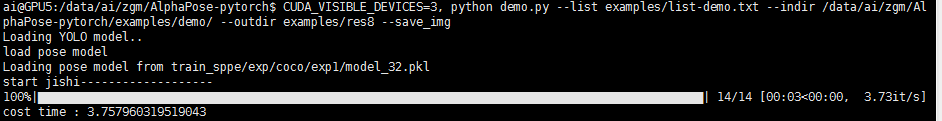


测试结果视频截图如下：



完整视频：<https://github.com/Erikfather/poetry-creation/blob/master/AlphaPose_test.avi>

**对图片进行测试：**



测试效果：



