Enhancing the Spatial Resolution of Stereo Images using Parallax Prior

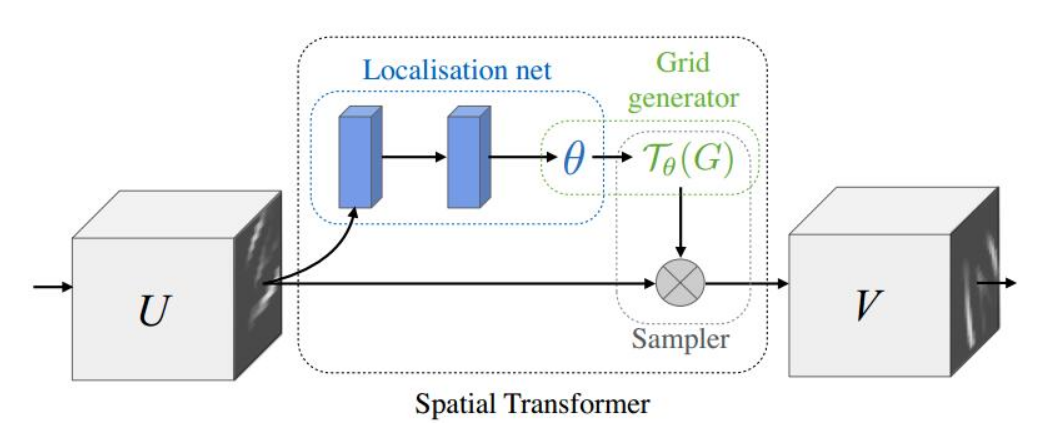
这篇论文的main idea 是使用DPN网络和STN网络结合，针对RGB+NIR融合过程出现匹配不准确的现象提出一种多模态融合方案。

## STN

首先看什么是STN网络。这里就不看原论文了，在CSDN上找了一些文章。

首先看STN的insight：

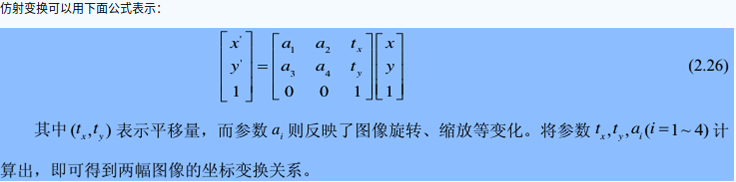
文章提出的STN的作用类似于传统的矫正的作用。比如人脸识别中，需要先对检测的图片进行关键点检测，然后使用关键点来进行对齐操作。**但是这样的一个过程是需要额外进行处理的。[进行什么样的额外处理？]**但是有了STN后，检测完的人脸，直接就可以做对齐操作。关键的一点就是这个矫正过程是可以进行梯度传导的。想象一下，人脸检测完了，直接使用ROI pooling取出人脸的feature map，输入STN就可以进行矫正，输出矫正后的人脸。后面还可以再接点卷积操作，直接就可以进行分类，人脸识别的训练。整个流程从理论上来说，都有梯度传导，理论上可以将检测+对齐+识别使用一个网络实现。当然实际操作中可能会有各种trick。



STN网络由Localisation Network ，Grid generator，Sampler，3个部分组成。

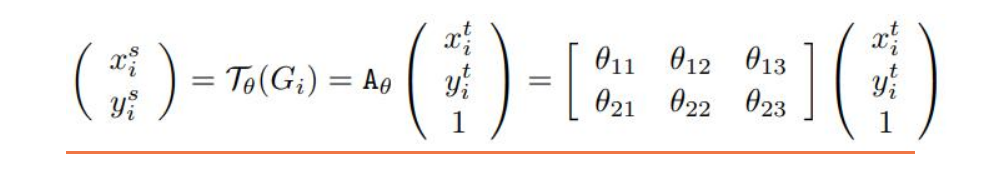
**Localisation Network：**

该网络就是一个简单的回归网络。将输入的图片进行几个卷积操作，然后全连接回归出6个角度值（假设是仿射变换），2\*3的矩阵。**【为什么是6个角度？】**

****

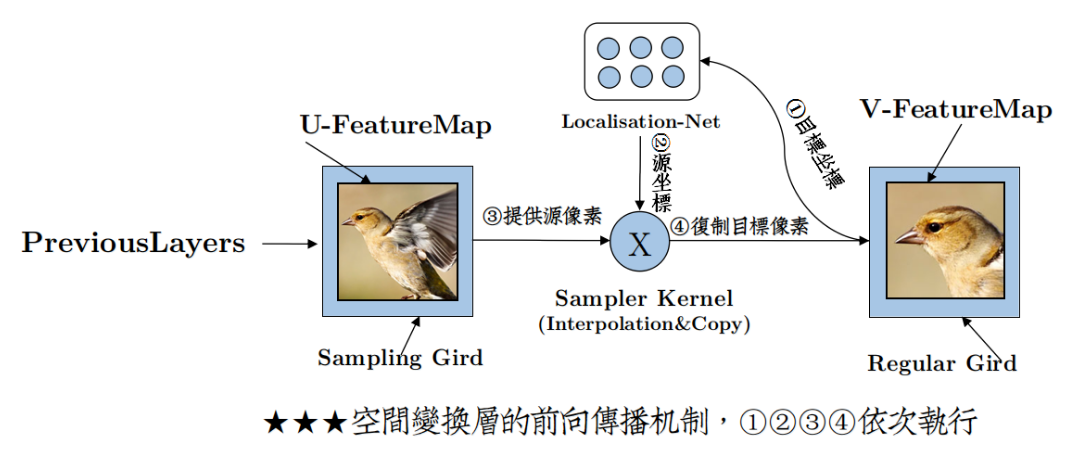
**Grid generator：**

网格生成器负责将V中的坐标位置，通过矩阵运算，计算出目标图V中的每个位置对应原图U中的坐标位置。即生成T(G)。



这里的Grid采样过程，对于二维仿射变换（旋转，平移，缩放）来说，就是简单的矩阵运算。

上式中，s代表原始图的坐标，t代表目标图的坐标。A为Localisation Network网络回归出的6个角度值。



整个Grid生成过程就是，首先你需要想象上图中V-FeatureMap中全是白色或者全是黑色，是没有像素信息的。也就是说V-FeatureMap还不存在，有的只是V-FeatureMap的坐标位置信息。然后将目标图V-FeatureMap中的比如（0，0）（0，1）......位置的坐标，与2\*3变换矩阵运算。就会生成出在原始图中对应的坐标信息，比如（5，0）（5，1）......。这样所有的目标图的坐标都经过这样的运算就会将每个坐标都产生一个与之对应的原图的坐标，即T(G)。然后通过T(G)和原始图U-FeatureMap的像素，将原始图中的像素复制到V-FeatureMap中，从而生成目标图的像素。

**【这步操作有点骚，为什么不生成从U到V的放射变换矩阵？】**

**Sampler的存在使得STN可微。**

**STN网络是一个CONV还是FC**

**这个网络会有几个隐层，这些隐层可以是全连接层，也可以是卷积层。**

**也就是说，STN是一种流程思想。**

这里没看明白。但后一层，需要做回归，因为要输出变换参数的值。

参考资料

[1]<https://blog.csdn.net/qq_39422642/article/details/78870629>，此博客中的连接很有用。

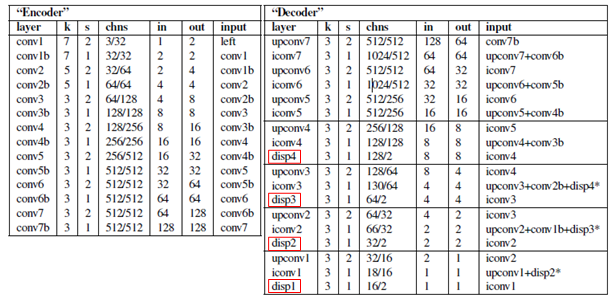
<https://blog.csdn.net/qq_14845119/article/details/79510714>

## DPN

**Baseline LR中的backward warp很有意思。有点STN中反step3的影子。**

**发现自己的一个概念一直用错了。右图的视差dr，应用到左图，可以重建出右图。**

原文中说，输出四个不同尺度（disp4~disp1）的视差图，它们前者是后者的两倍。尽管只有单张输入图片，该网络能够输出不同尺度上的两个视差图——left-to-right 和 right-to-left。



上面这张图是博客作者自己画的？反正没找到论文作者的版本。

网络输出两个视差图是毋庸置疑的，文章material awareness中示意图也说明了这点。

**为什么能够输出两个视差图呢？**

参考资料：

这一篇是DPN的一个baseline LR <http://yyliu.cn/post/750ec7f0.html>

下面两篇是翻译：

<http://www.yyliu.cn/post/c68cf6db.html>

<https://zhuanlan.zhihu.com/p/29528596>

这篇是概括性解读+code

<https://blog.csdn.net/lvhao92/article/details/76586101>

Deep Material-aware Cross-spectral Stereo Matching

Introduction

- 混合相机系统的align，如果使用Beam filter,会有长时间曝光造成blur.而使用stereo合一在align的同时获得depth

【相当于初始化depth用的是dpn单目方式】

- 作者让DPN只学深度，使用了symmetric的结构方式STN学习几何信息。

- 作者采集了数据集，并且自己label 了material segment进行train STN。

【看到这里作者的STN网络是不能借鉴了，猜测借鉴的部分还有wrap等】

Related Work

- 已有的交叉模态立体匹配方案只考虑了feature和region没有考虑material awareness

- 已有的非监督深度估计方案只考虑了RGB而没有考虑光谱，还有非漫发射。

【看到这里想到我们的方案是想让两路的图像尽可能接近

采用类DPN方案：可以得到disparity-->align-->enhance(这里enhance与求disp有冲突吗？没有。)求disp的时候可以算一个loss，这个loss主体是||L’-L||\_2+||R’-R||\_2；enhance的时候loss主体是？？？Enhance需要找1、jiangzhu的论文2、今年CVPR陈畅师兄介绍的zero-shot增强的论文。3、思考数据集怎么做？我们是没有label的。】

Simultaneous Disparity Estimation and Spectral Translation

- 文字贼有意思，simultaneously to respectively learn disparity and spectral translation

- 看到他用STN把RGB合成pseudo NIR，我在考虑是否应该使用CNN来把spectral image转化为灰度图像。那么这里为什么不直接用光谱响应曲线呢？考虑如下：1、合成后的灰度图像并不很好。【其实两幅图像的拍摄角度不同，光线也不会一样，所以导致spectral image合成后的图像也并不会完全一样。但是这个差别是否可以忽略不计呢?】

- reprojection error 重投影误差

- 出现了！出现了！使用STN进行transform(wrap)。。。吗【又看了一遍DPN后发现并不是，STN的作用：“在训练过程中，网络学习通过对相反的图片采样来生成图片。我们的成像模型，采用来自空间变形网络（spatial transformer network，STN），通过视差图对输入图像取样。STN采用双线性采样，它的输出像素是四个输入像素的权重相加 。与其他方法相比，双线性采样局部完全可微，无缝集成到我们的全卷积网络架构。意味着不需要单纯化或近似我们的损失函数。”】

Spectral Translation Network

- 包括局部滤波，白平衡，曝光校正。

- pseudo NIR 由RB白平衡G(\theta)和像素值F(\theta)加权得到。

- F(\theta)的权值由Filter Generating Network FGN学习得到。

- FGN的结构和DPN一样。此处为了不让STN学习到disp信息，所以使用了一个**left-right symmetric filtering kernels(symmetric CNN)**，即FGN。FGN可以平等地对待左右图的每一个像素并且不对输入产生shift【这里symmetric cnn是什么，为什么可以不学到disp信息？为啥对称了就学不到了？】【STN除了FGN以外还有什么组成】

Incorporating Material-aware Confidence into Disparity Prediction Network

- 提出一个问题：material awareness和disp prediction没有结合

- 解决方法：把material-aware confidence加到DPN loss中去。

- 方案1：把确定区域的disp传播到不确定性区域，（利用确定的disp求不确定的disp），为此，作者提出了一个新的confidence-weighted smoothing technique

- 方案2：改写DPN LOSS，增加material-specific alignment and smoothness losses

- 方案1+方案2.Done！

- smooth是一种预测未知区域disp的方法（这么讲好有道理）

- 方案1：简单来说就是对unreliable区域的disp进行类插值预测。

- 方案2：改造DPN loss貌似很复杂，不看了。

- 随后作者举例说明怎么使用方案1 2

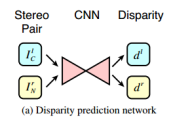
- 作者介绍了数据集

Experiment Result

- 使用了Deeplab net进行fine-tune，用于 material awareness

- 本文的对比对象是 CMA/ ANCC/ DASC /SIFT FLOW/

- ablation study

看到最后有个问题，本文中DPN的输入为什么会画成两个输入的形式？

并且输出为两个disp。这里的问题还是original 的DPN没有搞清楚。

-----------------------------------------------

20180917再次阅读。

主要阅读method部分。追求技术细节。

首先，DPN的作用是求视差，这里没有问题。关于其约束，使用了重投影误差，查了一下重投影误差并没太仔细看，觉得这里可以先简单理解为左右一致性。其loss设置的比较复杂，但感觉基本就第一项在起作用。

考虑其训练过程和测试过程。