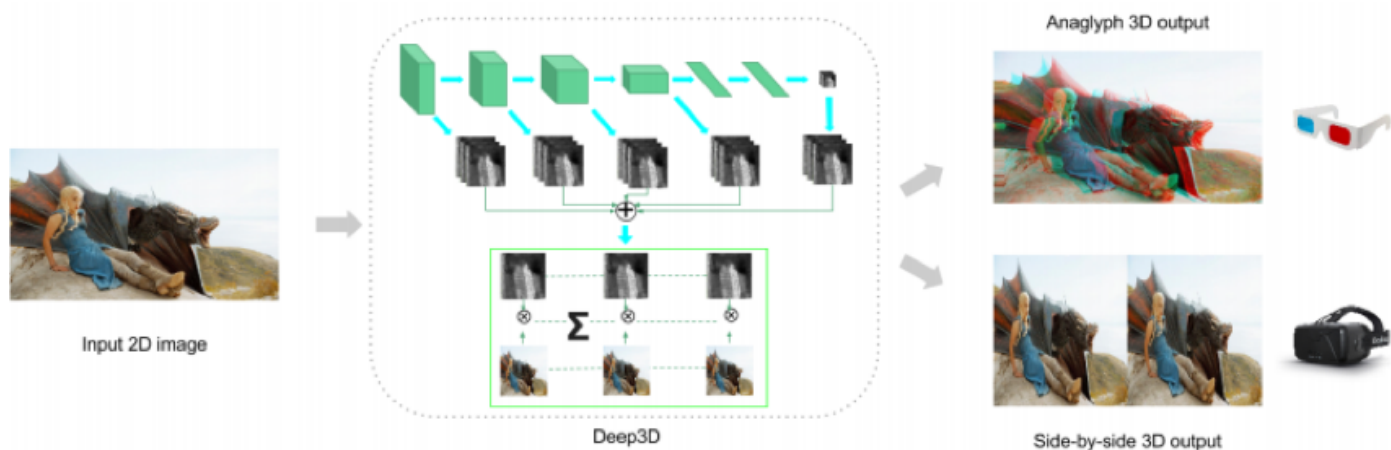


# 基于深度学习的新视角图像生成方法

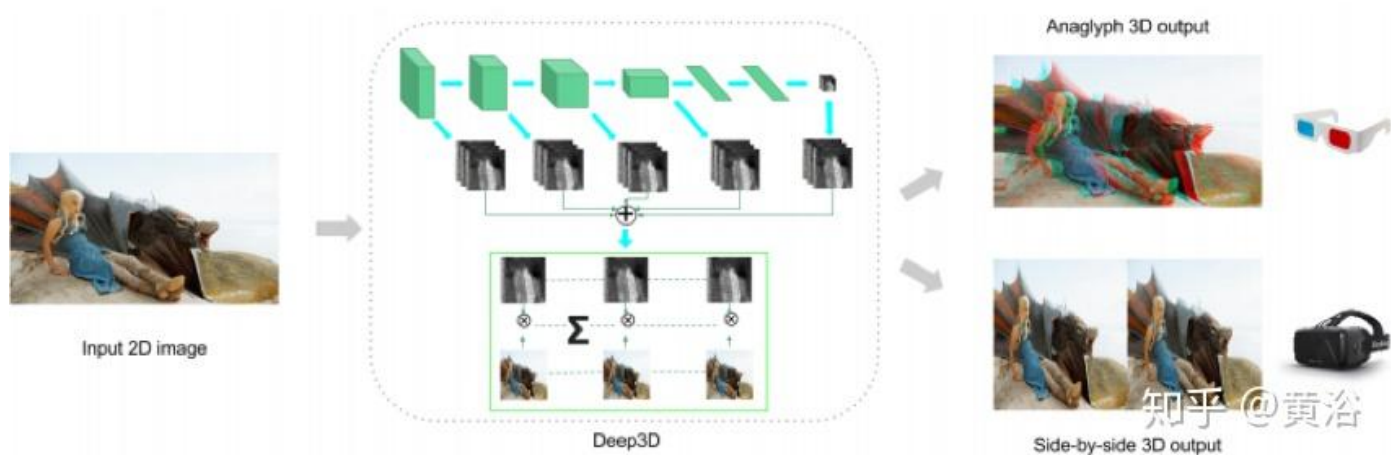


基于已知的不同视角的3-D几何关系，可以从相邻视角图像去“内插”或者“扩展”新视角图像。以前，这称之为基于图像的绘制技术（IBR），机器学习方法也曾发挥过重要作用，这里介绍一下深度学习的方法。

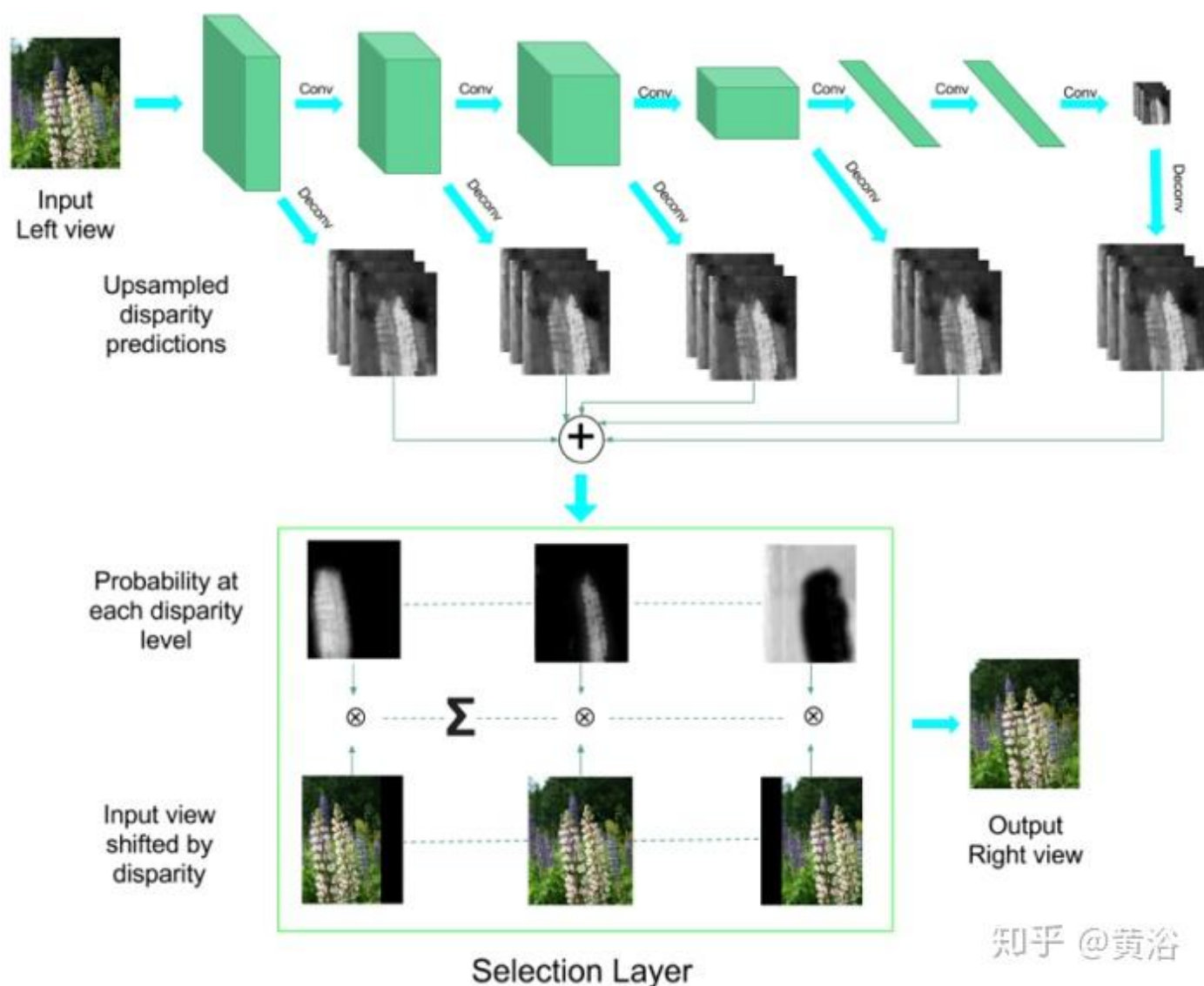
## • Deep3D

这种方法用深度神经网络自动将2D视频和图像转换为立体3D格式。它直接在从现有3D电影中提取的立体图像对上进行端到端训练。这种训练方案可以利用数量级更多的数据并显著提高性能。

下图是Deep3D的工作示意图：在内部，Deep3D网络估计概率视差图，该视差图进入可微分的基于深度图像的绘制（depth image-based rendering, DIBR）层以产生正确的新视图。



下图是Deep3D架构图。模型结合了多层的信息，进行端到端的训练，以便从左视图直接生成正确的右视图。基础网络预测类似视差的概率图，然后由选择层（selection layer）以差分方式对DIBR建模，其中允许隐含的修补（inpainting）操作。



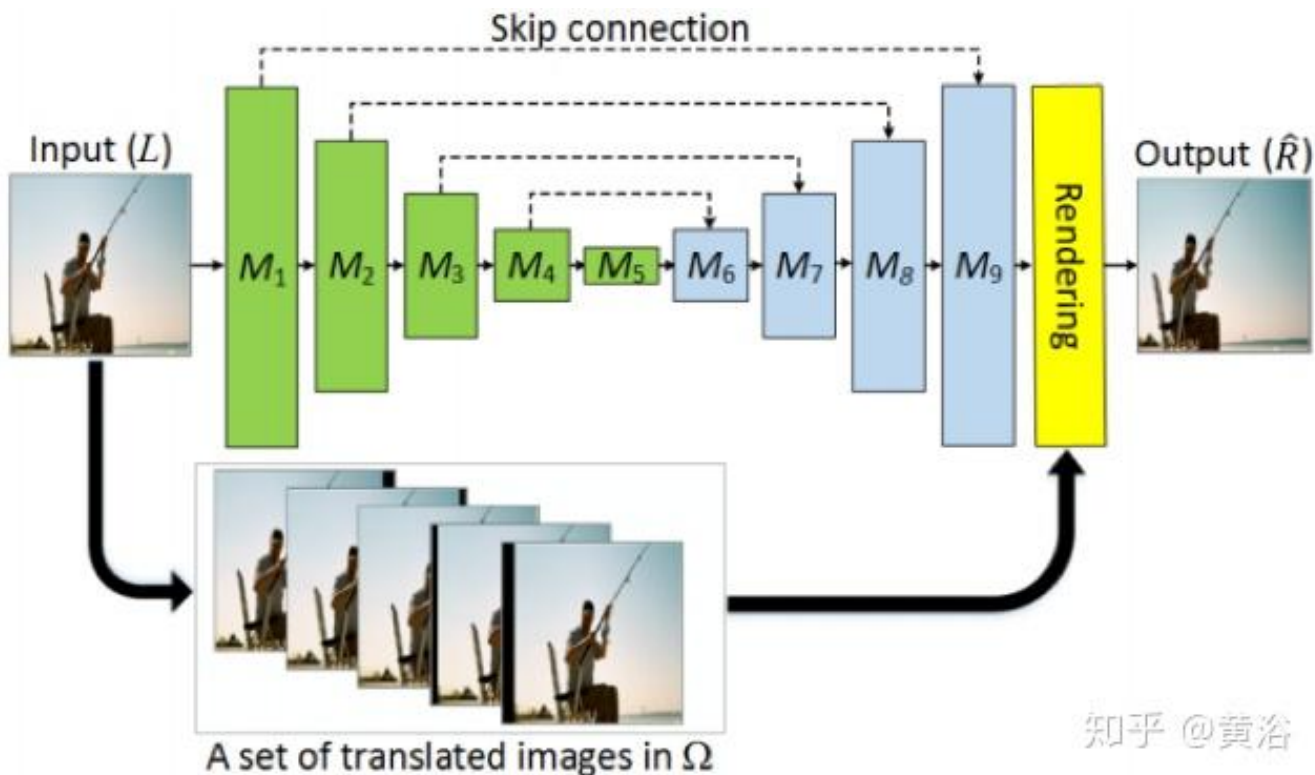
## • SIVG with Fully Convolutional Networks

基于单图像的视图生成（Single-image-based view generation, SIVG）采用全卷积网络（FCN）。设计两种FCN架构。第一个是基于FCN和名为DeepViewren的视图绘制网络的组合。第二个由亮度（luminance）和色度

（chrominance）信号的解耦网络组成，由DeepViewdec表示。采用了一个2M立体图像的大型训练数据集。

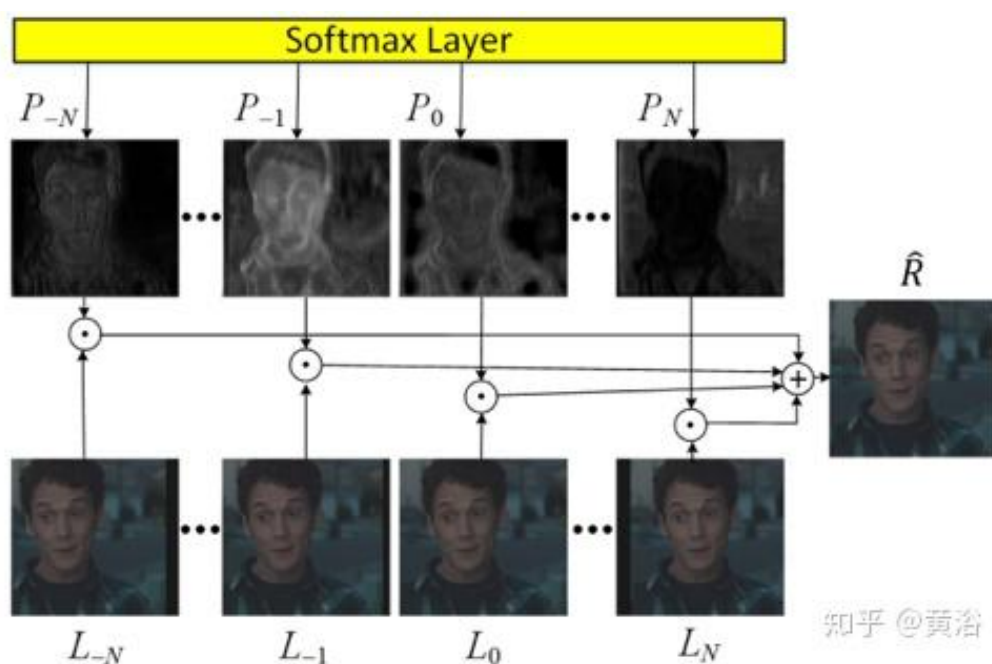
DeepViewren处理速度很快，而DeepViewdec具有更高的精度。

下图是DeepViewgen架构。编码、解码和渲染网络分别以绿色、蓝色和黄色显示。编码网络从输入图像中提取低、中和高级特征，并将它们传送到解码网络。在解码之后，渲染网络生成概率视差图并估计右图像。这里，使用一组平移图像（参见下面的绘制网络）。



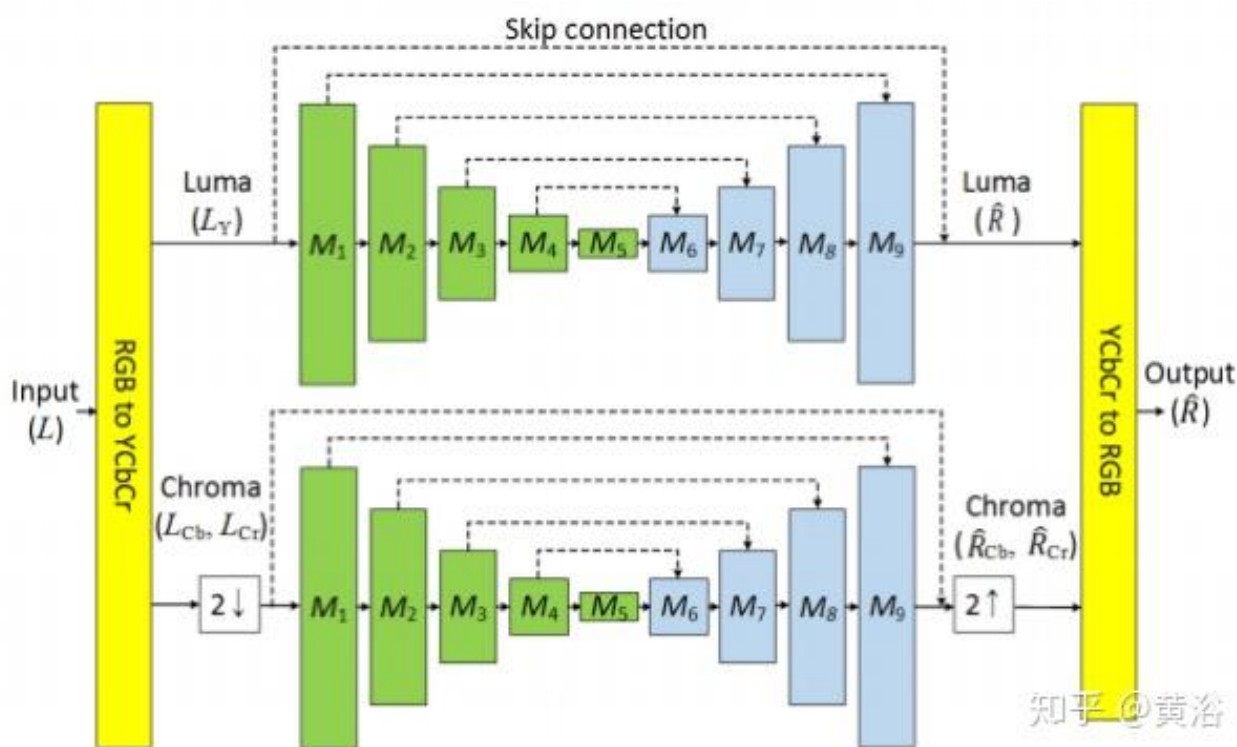
知乎 @黄浴

下图是绘制网络架构，softmax层将解码网络的输出归一化为信道上的概率值 ( $P_i, i \in \Omega$ )。这里，信道的数量与视差范围  $\Omega = \{-N, -N + 1, \dots, 0, 1, \dots, N\}$  的视差数量相同。最终右视图图像  $R$  通过  $P$  与其相应平移的左图像  $L$  之间的逐像素相乘来合成。



知乎 @黄浴

下图是DeepViewdec架构。 DeepViewdec由两个具有相同架构的解耦网络组成，即亮度（Y）和色度（Cb, Cr）网络。每个网络都是单独培训。每个网络中的绿色、蓝色分别定义编码和解码网络，而黄色表示RGB和YCbCr之间的颜色转换。 RGB图像输入被转换为Y, Cb和Cr图像，其中Cb和Cr图像以2倍进一步缩小。编码网络从输入图像中提取低、中和高级特征并将它们传输到解码网络。解码网络的推断图像被逆变变为输出RGB图像。 注意，亮度网络仅用Y通道图像训练，而色度网络用Cb和Cr通道图像训练。

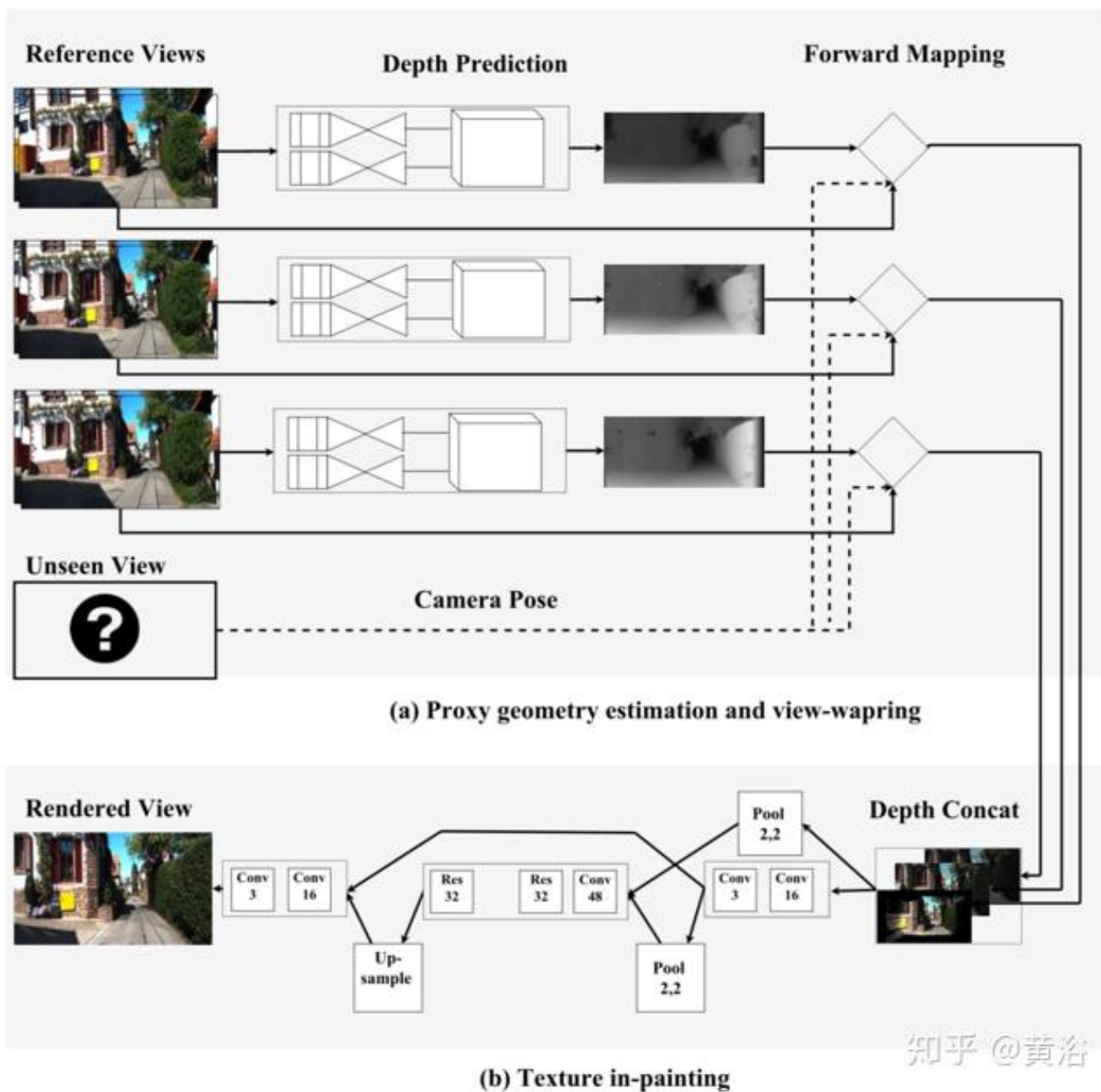


## • Deep Stereo Vision

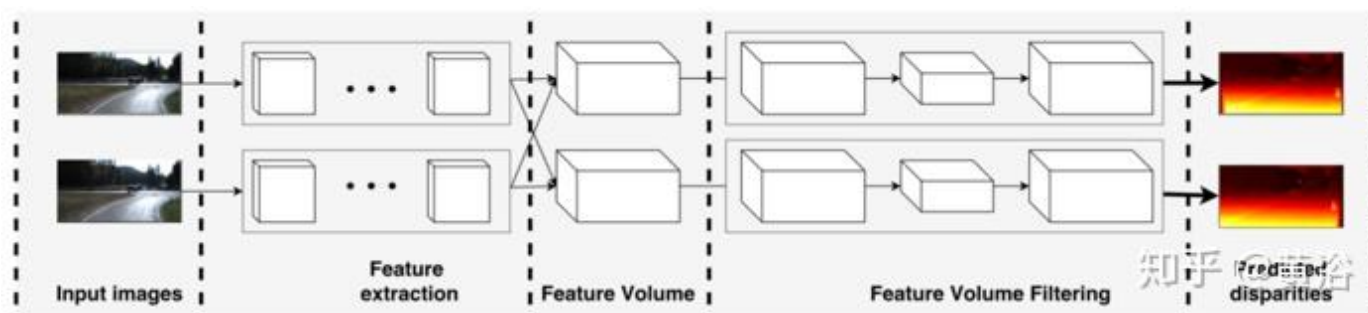
大基线新视图合成问题远未被“解决”。这是一种基于立体视觉和CNN的方法，它将问题分解为两个子任务：视角相关的几何估计和纹理修复。这两项任务都是结构化的预测问题，可以通过CNN有效地学习。

下图是该视图合成方法的示意图。在第一阶段（a）中，首先使用无监督的立体深度预测网络从输入的参考立体图像对估计致密深度图。估计的深度图用于通过前向映射将输入视图投影到目标视图，如（b）所示，经典的DIBR技术，而输出新视图是在前向映射视图上应用纹理修复渲染而得。





而下图是深度预测网络的架构图，由以下阶段组成：特征提取、特征容积聚合和特征容积滤波。该网络以无人监督的方式进行训练，不需要基础事实（GT）数据。



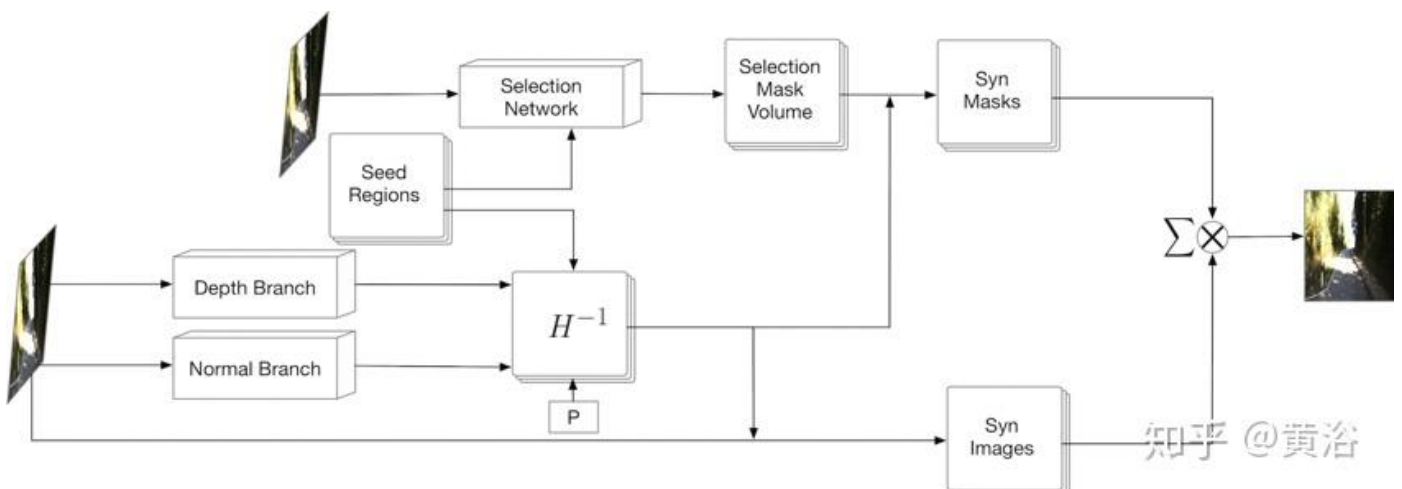
## • Geometry-aware Deep Network

该方法主张利用场景的三维几何来合成新视图。通过固定数量的平面逼近现实世界的场景，并学习预测一组单应性

(homographies) 及其对应的区域掩码，这样可以将输入图像转换为新视图。为此，提出一个对区域觉察的几何变换网络。

下图即是这个区域觉察的几何变换网络结构。为了解决单图像新视图合成问题，提出一个由两个子网组成的几何觉察的深度学习结构。第一个子网络，给定输入图像，能预测逐像素的深度和法线图；然后，将这些预测与从图像获得的分割掩码和所需的相对姿态结合一起使用，以产生固定数量的单应性变换，之后用于产生变形图像。第二子网预测逐像素的选择图

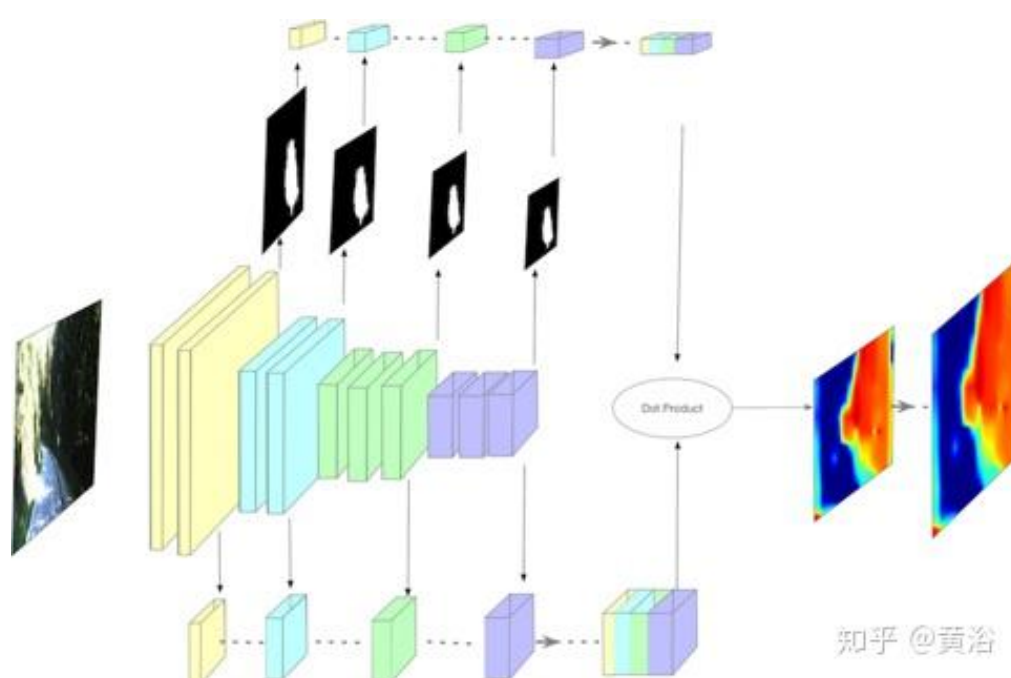
(selection map)，其将每个输入像素与一个单应性变换相关联。这些选择图由各自的单应性变换变形，据此组合变形图像来最终生成新视图。



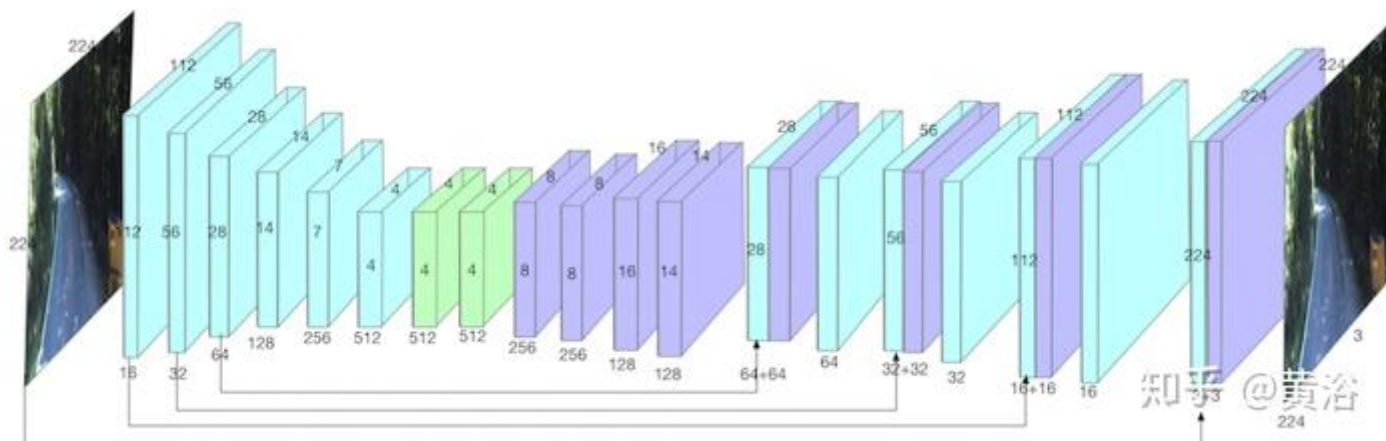


而这个选择网络结构见下图：不是使用硬分割掩码来组合候选合成图像，而是训练网络生成一组软选择掩码。网络结构遵循VGG16的前4个块的网络结构。根据种子掩码对相应的4个特征图做最大池化（max pooling），并在超列特征

（hypercolumn feature）中连接得到的4个特征向量。然后，将这个超列特征与低分辨率的级联完整特征图卷积，从而生成一个上采样到原始图像大小的全局热图。



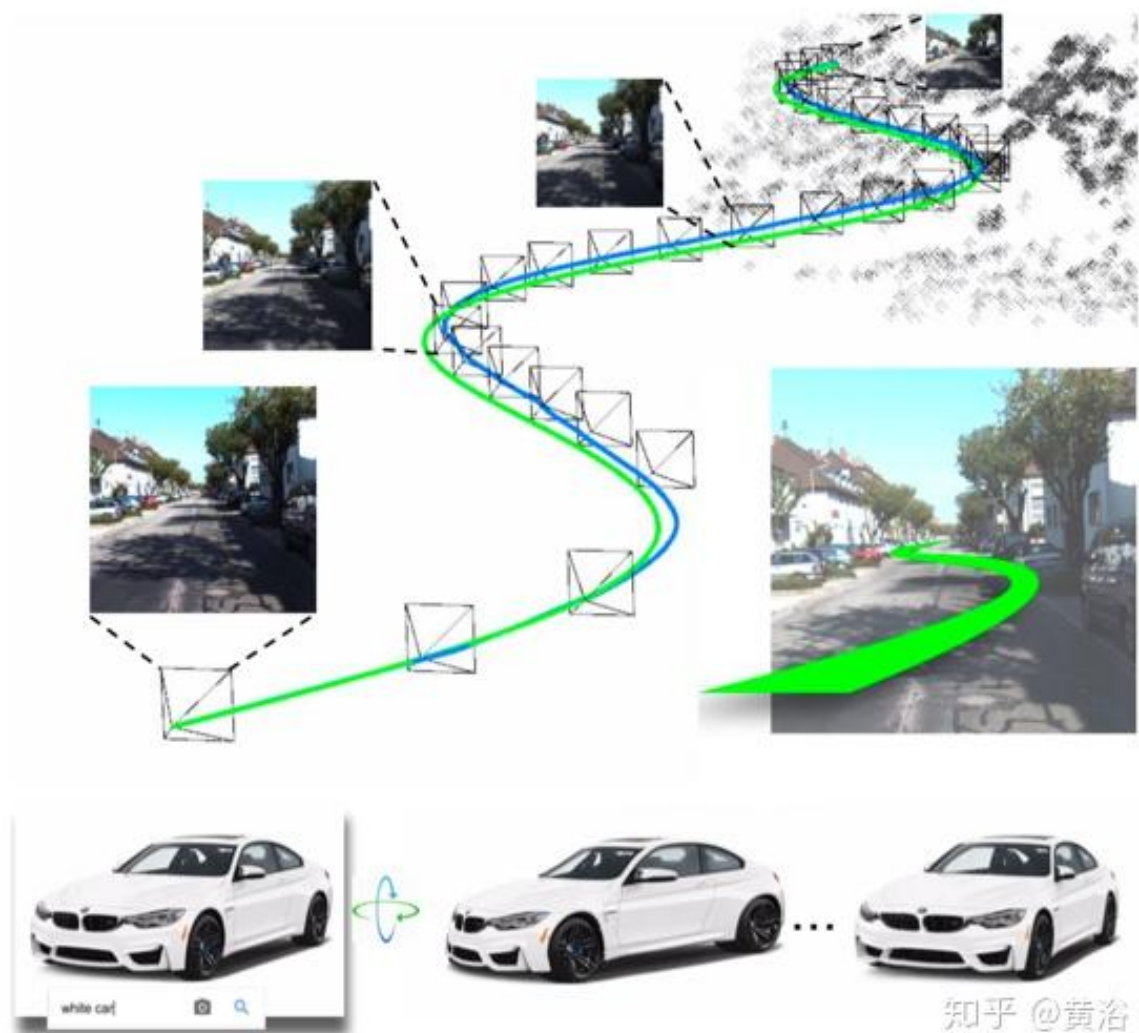
区域觉察的几何变换网络产生了一种新视图图像，可以保持场景的局部几何结构。虽然几何变换可以合成同时出现在输入图和新视图中的区域，但它无法处理仅出现在新视图中的区域，即在输入视图中的隐藏区域。为此，采用一种如下图所示的编码器 - 解码器细化网络（refinement network）



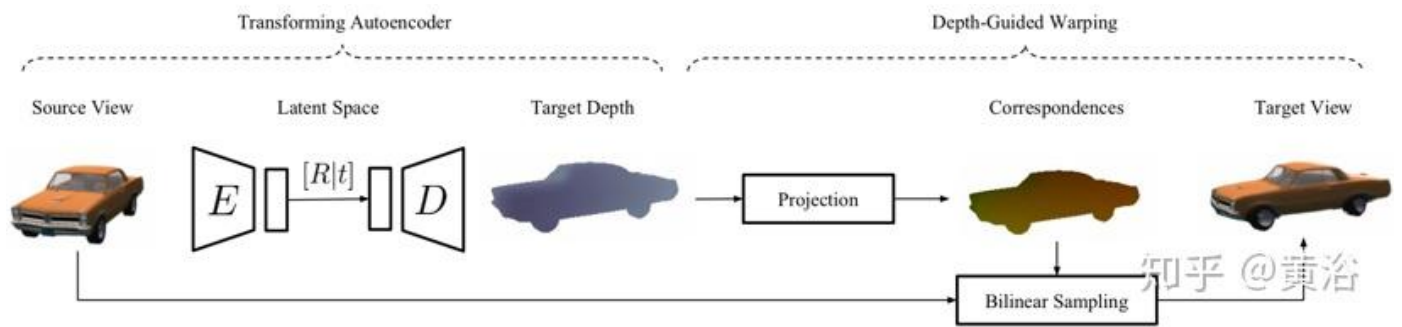
## • NVS Machines

这是一种可以学习如何合成高质量3D场景新视图的方法，同时提供对6-DOF视点的细粒度加精确地控制。该方法是自我监督方法，只需要2D图像和相关的视图变换进行训练。网络架构利用变换自动编码器（transforming auto-encoder）结合深度引导的变形程序来预测几何精确的新视图。利用几何约束，无需通过深度或光流图，直接监督学习。如果在源视图中对象被大部分遮挡，则使用纯基于学习的先验来预测露出像素的值。网络还预测了逐像素掩码，用于融合深度引导的基于像素的预测。最后的结果图像反映了所需的6-DOF变换，细节也得以保留。

如图是NVS机示意图：给定单一源视图，生成具有细粒度控制的精确新视图。上图：在给定单个输入源和用户指定的3D轨迹情况下生成合成视图。绿色：GT轨迹；从合成视图中恢复的蓝色轨迹（长度为5m）。底部：来自互联网的高分辨率图像用于合成新的细粒度视图；整个模型训练仅使用合成数据。

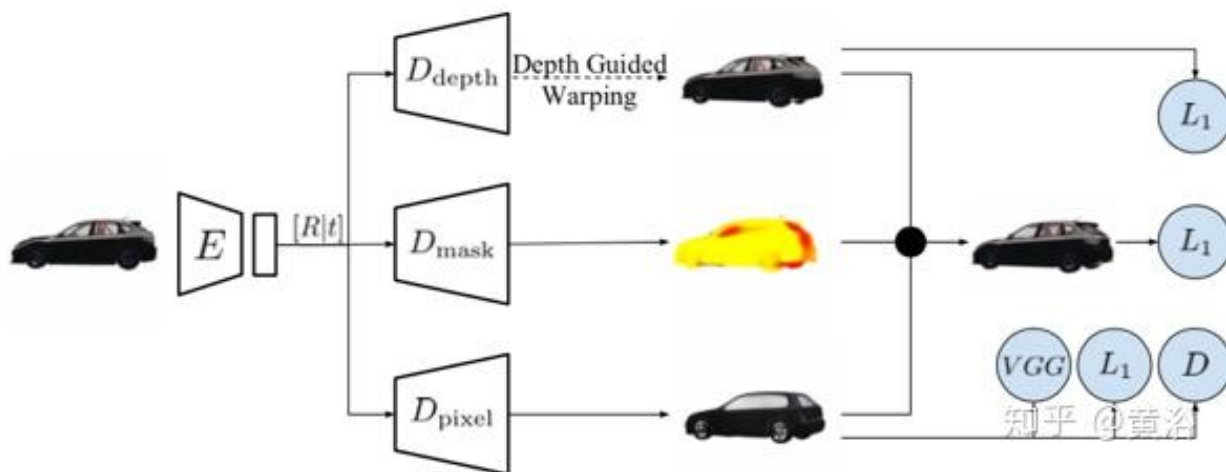


如下图给出合成方法的基本结构图。包括3D变换自动编码器（T-AE），自监督深度图预测，深度图投影和表观变形。深度觉察的分支提供细粒度的相机控制，例如允许围绕对象平滑地旋转相机，而不是固定的视角。此外，它可以产生高度准确的预测，特别是对于只有轻度暴露的情况。然而，对于最初隐藏许多像素的情况，图像表观基于学习的先验是有效的。为了处理这两种情况，网络中有额外的分支，它们将变换后的潜在表示直接解码到目标视图（图像分支）和逐像素的掩码中。通过深度和基于图像的预测之间的加权平均来获得最终预测，其中掩码提供权重（参见下面的加权平均图）。

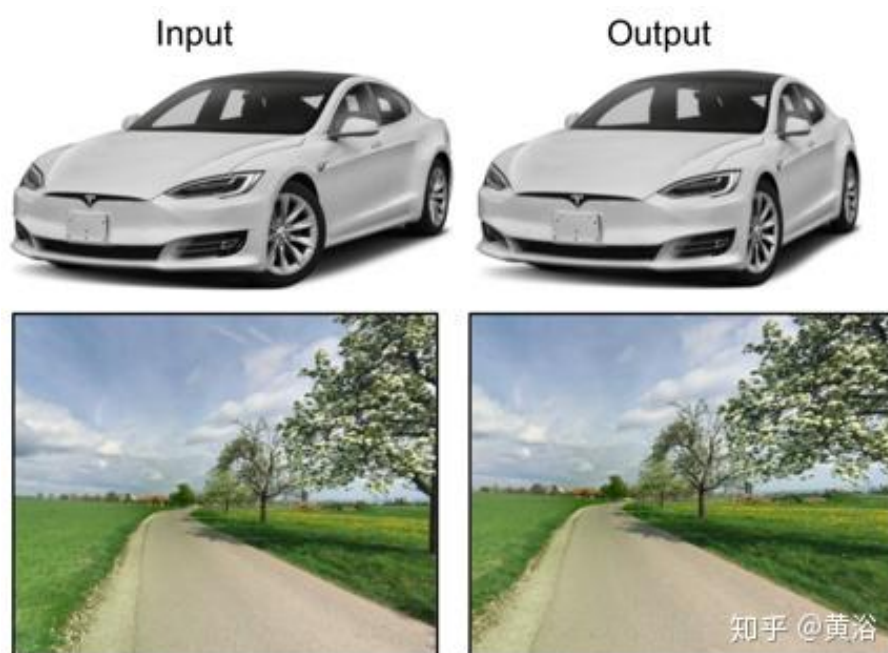


首先将源视图编码为潜代码 ( $z = E\theta_e(l_s)$ )。整个学习方案对这个潜代码 $z$ 是鼓励的，使它在3D度量空间中有意义。编码后，将源和目标之间的所需转换应用于潜代码。变换后的代码 ( $z_T = T_{s \rightarrow t}(z)$ ) 由神经网络解码，预测从目标视点观察到的深度图 $D_t$ 。然后基于已知的相机内参数矩阵 $K$ 和外参数 $T_{s \rightarrow t}$ 将 $D_t$ 投射到源视图，产生目标视图和源视图之间的致密匹配，编码为致密的后向光流图 (backward flow map)  $C_{t \rightarrow s}$ 。然后用该光流图将源视图逐像素地变形到目标视图中。整个网络采用端到端训练，唯一的监督是目标视图和基础事实 (GT) 图像之间的L1重建损失。

如图是NVS机 (加权平均) 一览图。NVS机架构通过加权平均结合了基于深度 (顶部) 和直接像素预测 (底部)。权重在逐像素掩码 (中间) 中编码。网络以自我监督的方式进行端到端的训练。



如下是在两个数据库ShapeNet和Kitti分别取图像做源视图的合成结果例子（上：ShapeNet数据库，下：Kitti数据库）。



## 参考文献

1. J Xie et al., "Deep3D: Fully Automatic 2D-to-3D Video Conversion with Deep CNN", arXiv 1604.03650, 2016

2. S Bae et al., “Efficient and Scalable View Generation from a Single Image using Fully Convolutional Networks”, arXiv 1705.03737, 2017
  3. T Habtegebrial et al., “Fast View Synthesis with Deep Stereo Vision”, arXiv 1804.09690 , 2018
  4. M Liu, X He, M Salmann, “Geometry-aware Deep Network for Single-Image Novel View Synthesis”, CVPR 2018
  5. X Chen, J Song, O Hilliges, “NVS Machines: Learning Novel View Synthesis with Fine-grained View Control”, arXiv 1901.01880, 2019
- 

这种新视角合成的方法在自动驾驶模拟仿真中可以得到应用。