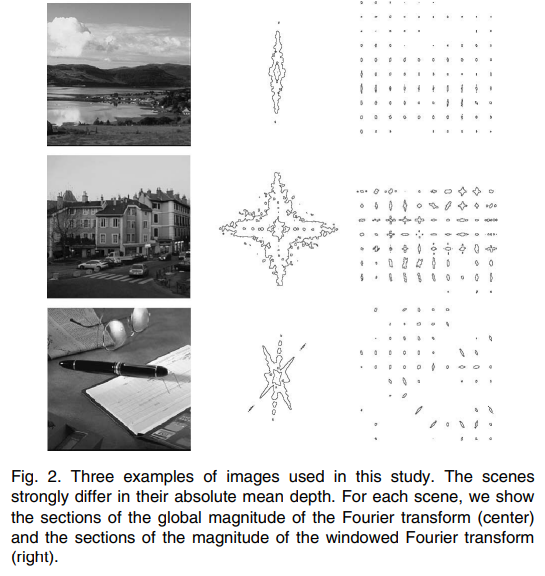
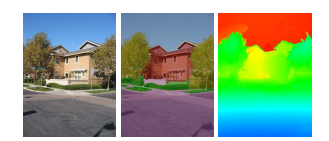
總體而言我們使用兩階段的偵測模型來解決生成自然散景的功能，首先利用深度估計模型來做物體深度的估計。

單鏡頭的深度估計從以前到現在有許多相關的論文。

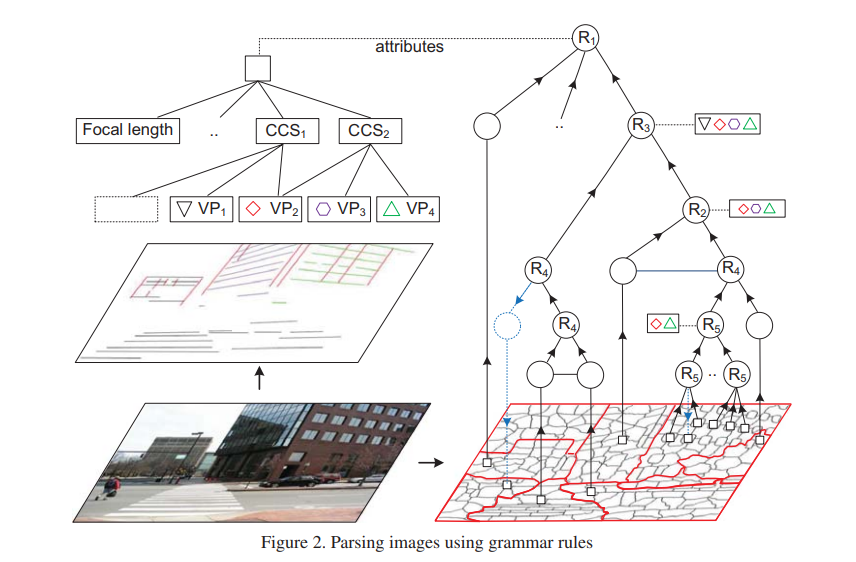
像是 ”Depth estimation from image structure” 是利用影像的結構來估計影像深度，論文中提到利用傅立葉轉換方式得到頻譜圖後，再利用其數值利用迴歸方式去預測整張影像深度



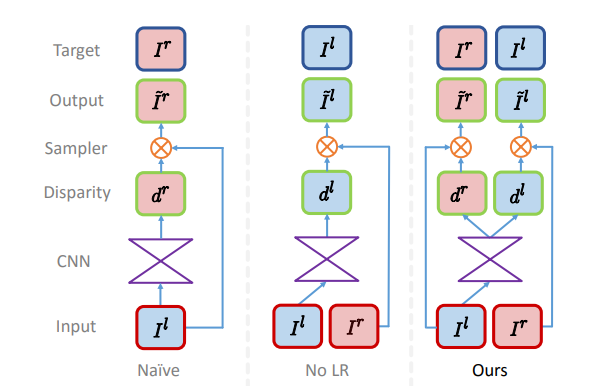
“Single image depth estimation from predicted semantic labels” 一文是利用影像的語意資訊，如同建築物、樹、道路等並計算他們的相對位置關係來求得其個別深度關係。



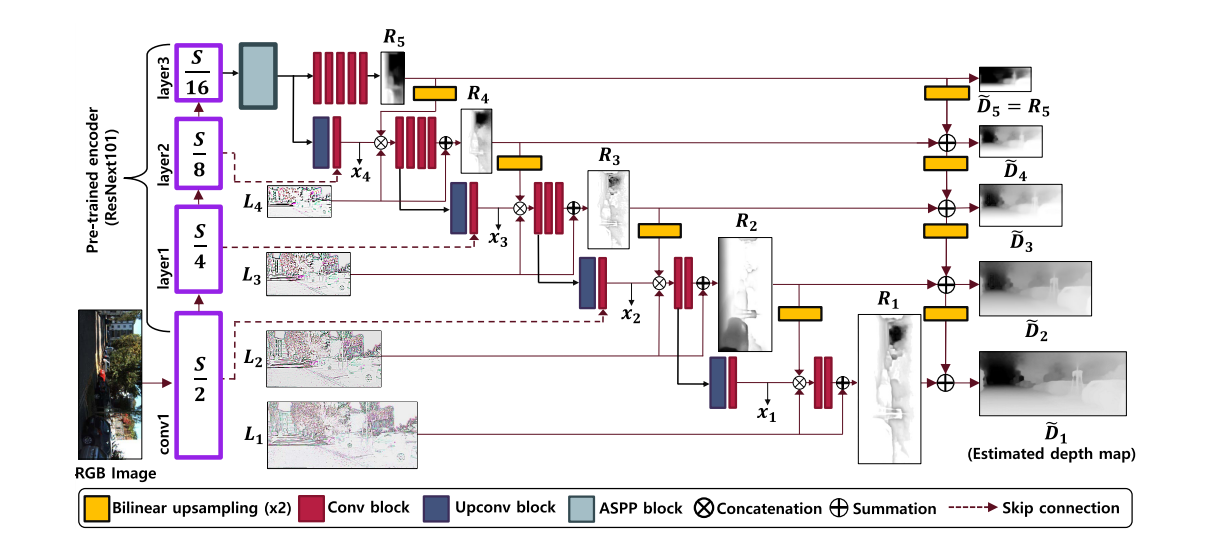
“Single-view 3d scene parsing by attributed grammar” 是利用 superpixel 資訊去分割出個物件，並且以其正交、相對位置等關係推論出物件位置。



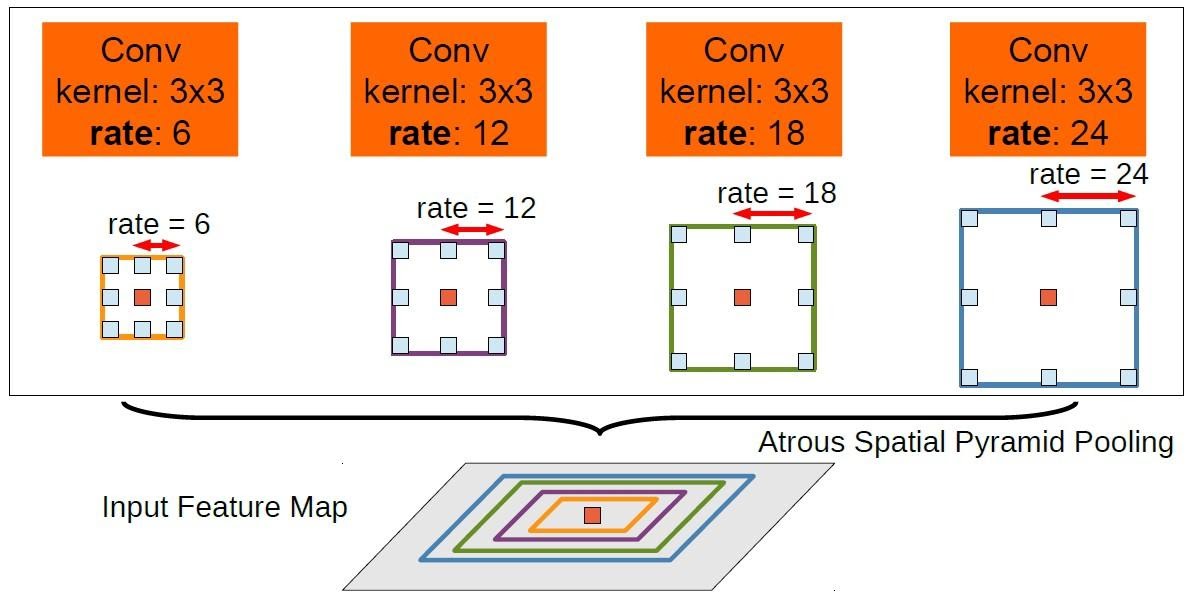
“Unsupervised monocular depth estimation with left-right consistency” 利用生成神經網路(Generative Adversarial Network) 以單目影像去生成另一側的影像以得到雙目的影像再用其雙目影像作計算以得到所需深度資訊，也是現代許多單目深度估計使用的方式。



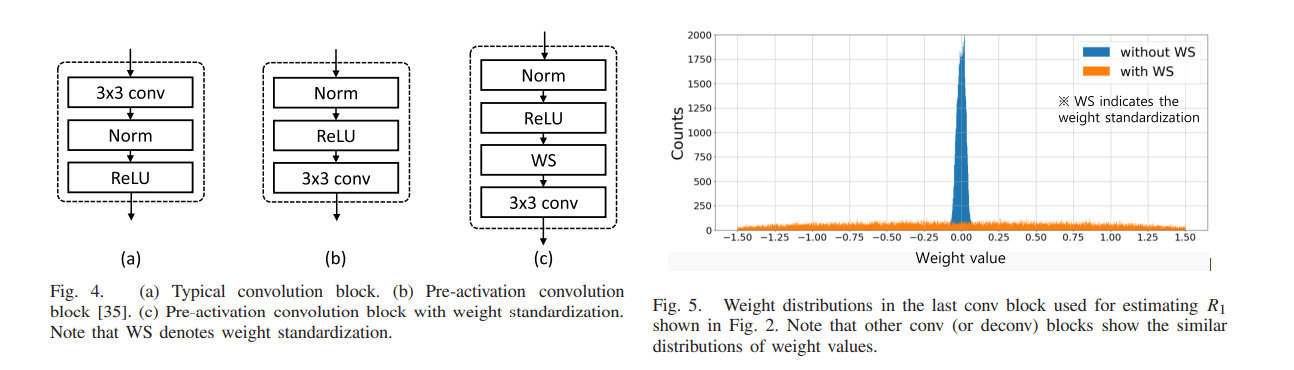
我們在此次Implement選用 “Monocular Depth Estimation by Laplacian Pyramid-Based Depth Residuals”作為我們的 Model 架構，是因為不管是 KITTI Eigen split 資料集或者 NYUv2 資料集，光資料集大小皆為200GB 以上，在整體訓練時間及設備要求算是單目深度估計的一種壁壘，也是因龐大的資料量讓單目深度估計的任務看過大部分情形，於是乎在真實世界中也較易使用上。



而稍微講解一下一下此次的模型架構，我們可以看到首先利用ResNext101作為特徵編碼器，以及利用原始影像利用Bilinear插值縮小再放大之後將相同大小的影像做相減可得知在過程中會遺失的資訊，在此以L1至L4做為表示，以利之後在up convolution中還原時能有效將邊界良好還原。然而Decoder部分則會同時使用Feature extractor及細部邊緣資訊。如同上圖所示Decoder會利用不同維度大小的資訊來層層做影像深度預測，並且結合前一層的解果做Upsampling以及邊緣資訊，經過多層採樣之後得到深度結果如架構裡面得圖所示。



然而架構圖所提到的 ASPP模塊則是以混合多感受域的，能有效利用物體周遭物件的資訊。



在論文中也有提到利用權重標準化的技巧來增加訓練時梯度的改變，避免使得梯度消失導致梯度一直無法做更新，導致結果不佳。

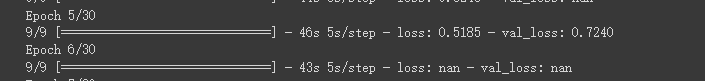
然而在訓練時利用以底下的Loss function去做更新約訓練50個週期

is the predicted result y is the ground truth , is the balance factor

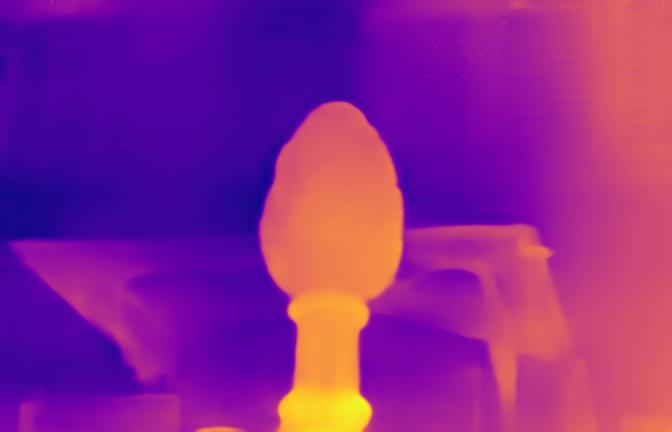
means RMSE of

is gradient loss of y and y\* (gradient y we can write as and )

前30個週期利用實際及預測距離取對數做 RMSE 來達成大致上深度預設正確，剩下部分則利用RMSE及gradient loss(分別求垂直及水平方向)並利用alpha及beta來做權重分配以達成對邊緣部分的精修。

我們在這也有嘗試利用較小資料集做訓練，但因可能pretrained weight不夠好或者warm up 參數設置不佳在訓練沒幾個週期後就超出整個預測範圍。

以下是我們利用深度估計模型所得到的結果



可以從結果明顯看到可有效識別不同物件深度。

在散景製造上面我們參考了 “AIM 2020 Challenge on Rendering Realistic Bokeh”，其競賽目的為創造更自然的散景效果，在其中許多團隊都利用End to end 的深度學習模型去製造出更自然的散景效果。然而相對於固定一個聚焦點，我們更有興趣去自己手動選擇聚焦點。於是乎我們因此去下載了EBB! bokeh dataset並且仔細觀察DSLR大光圈鏡頭下拍出的散景效果，於是我們想利用多層次的高斯模糊去模擬散景效果。



由左至右是我們套用不同深度的高斯模糊來製造出的散景效果。

我們會依照深度估計出來的數值來動態設定其模糊區間我們設其為 ()，接下來我們利用對焦點以及各點距離分割出不同的區域，公式如下所示：

(Depth of focus point), (Depth of the current point)), (Depth of different bokeh strength)

Region 1:

Region 2:

Region 3:

Region 4:

Region 5:

並將各區域遠近跟不同的高斯模糊影像做相乘得到最後多層次景深的結果，如同下圖所示：



可看到離水瓶最遠處的角落有最模糊的結果，相對的其他位置也有或多或少的模糊感，有效製造出自然的散景效果。