# Unsupervised Learning of depth and ego-motion from video using 3D geometric constraints.

## Summary

1. The loss for training is :

$$L = \sum_s \alpha L_{rec}^s + \beta L_{3D}^s + \gamma L_{sim}^s + \omega L_{SSIM}^s$$

2. s denotes the different scales of the input image.

3. $L_{rec}^s = \sum_{ij} \| (X_t^{ij} - \hat{X}_t^{ij}) M_t^{ij} \|$

   3.1 this is the "photometric consistency loss"

   3.2 $X_t$ is image at time $t$

   3.3 $\hat{X}_t$ is reconstructed image at time $t$ by warping $X_{t-1}$ using $D_t, T_t$.

   3.4 $M_t$ is a mask

4. Principled mask $M$ : they compute $M$ analytically from the depth and ego motion estimate.

5. $L_{3D} = \| T_t' - I \|_1 + \| r_t \|_1$

6. $L_{3D}$ is a 3D point cloud alignment loss

7. Given two point clouds $\hat{Q}_t$ and $Q_t$, they use Iterative Closest Point (ICP) to compute a transformation $T'$ that minimizes the distance b/t transformed point in $\hat{Q}_t$ & their corresponding points in $Q_t$.

8. $r_t$ is the residual, which reflects the residual distance between corresponding points after ICP's distance minimizing transform $T'$ has been applied.

9. $L_{sm}$ is a depth smoothness loss:

$$L_{sm} = \sum_{ij} \| \partial_x D^{ij} \| e^{-\| \partial_x X^{ij} \|} + \| \partial_y D^{ij} \| e^{-\| \partial_y X^{ij} \|}$$

10. By considering the gradients of the image, this loss function allows for sharp changes in depth at pixel coordinates where there are sharp changes in the image.

11. $L_{SSIM}$ is a structured similarity loss, usually used to evaluate quality of image predictions.

12. $$L_{SSIM} = \sum_{ij} \left[ 1 - SSIM(\hat{X}_t^{ij}, X_t^{ij}) \right] M_t^{ij}.$$

$\mu, \sigma$: local mean & variance

$$SSIM(x,y) = \frac{(2\mu_x \mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x + \sigma_y + c_2)}$$