Optimization to the Unification & Unified Focal Loss of UniMVSNet

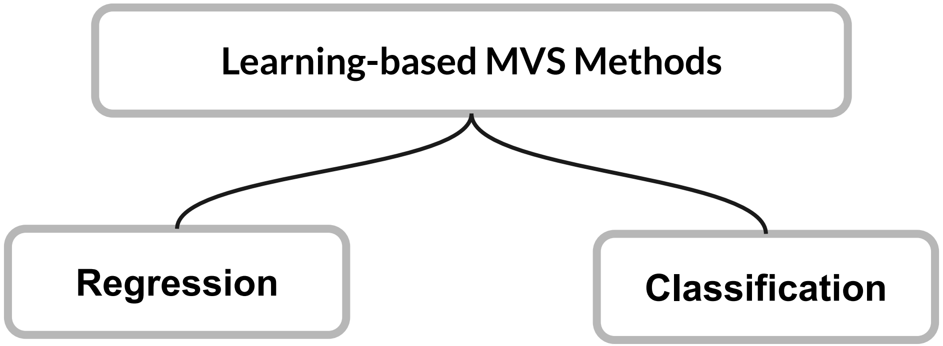
Based on CVPR 2022:
Rethinking Depth Estimation for Multi-View Stereo: A Unified Representation

Group9 - 311605004 劉子齊 & 311553014 張亭萱

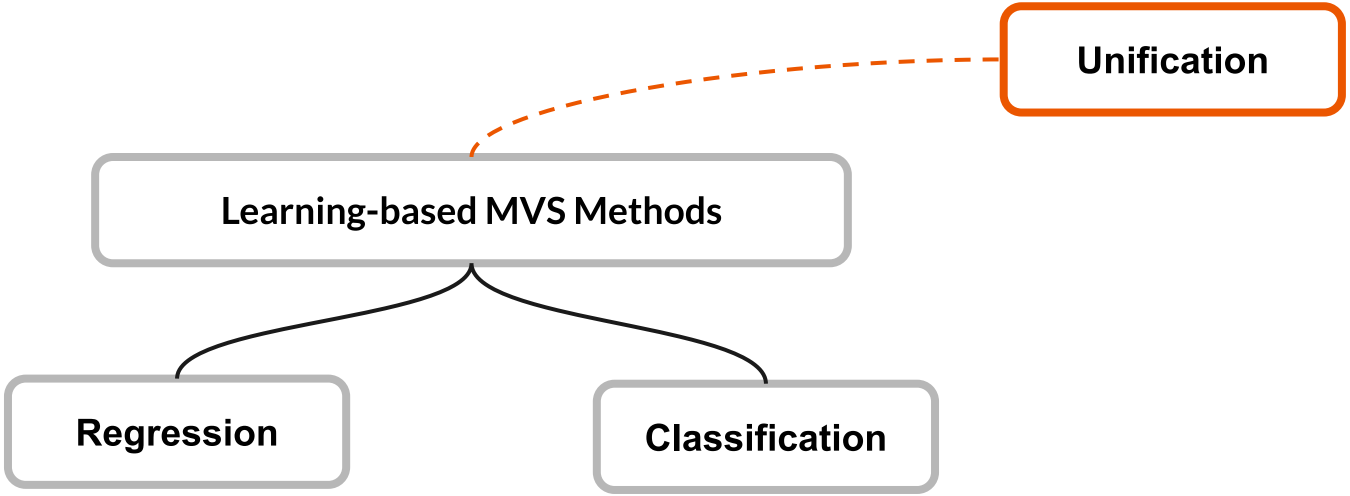
Outline

- Introduction
- Methodology
- Proposed Methods
- Experiment Result
- Conclusion
- Q&A

Introduction



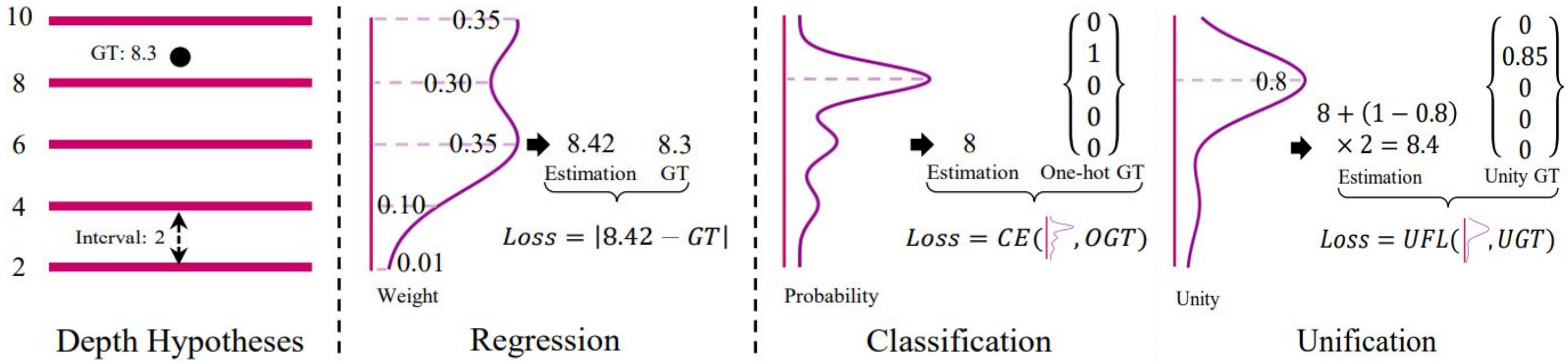
Introduction



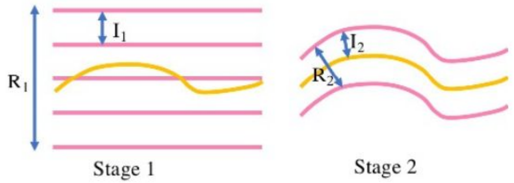
Introduction

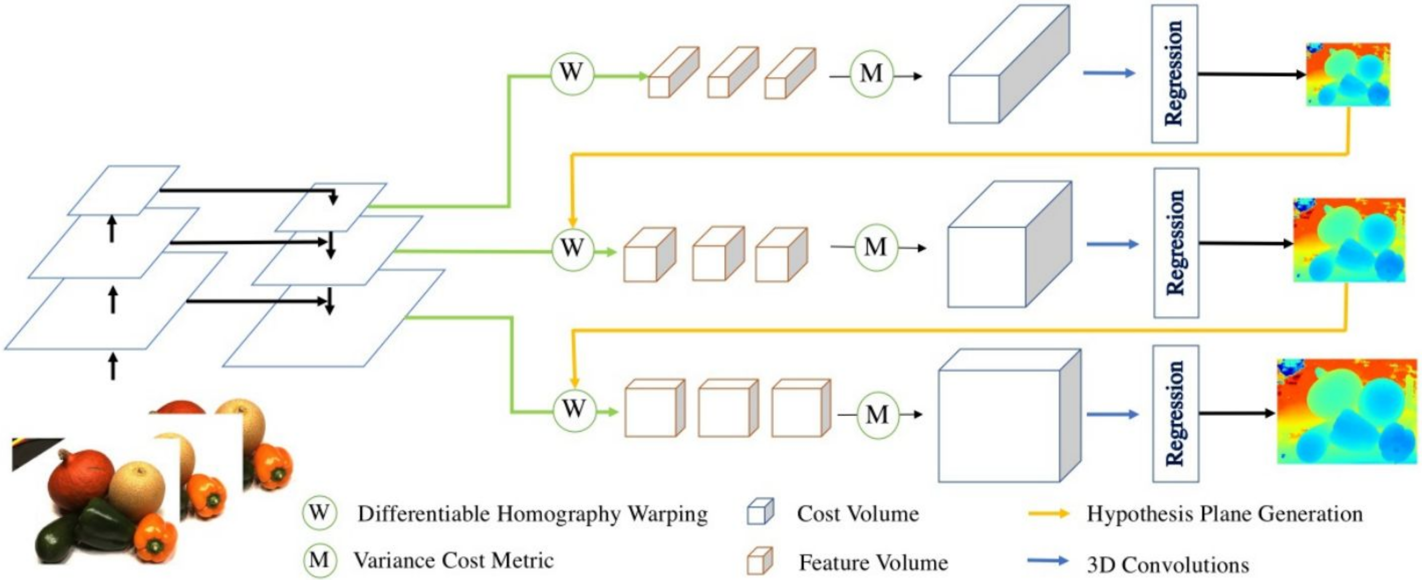
Unification ≈ Classification + Regression

Introduction



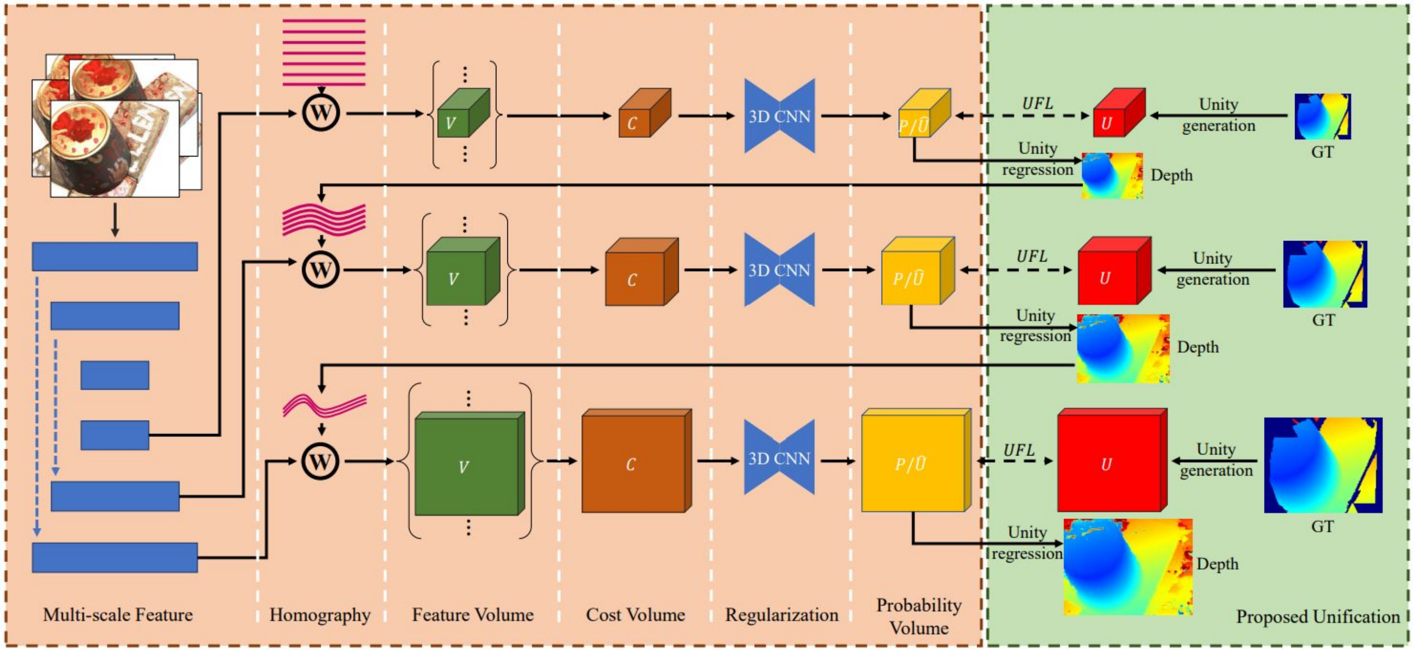
Related Works - Cascade MVSNet



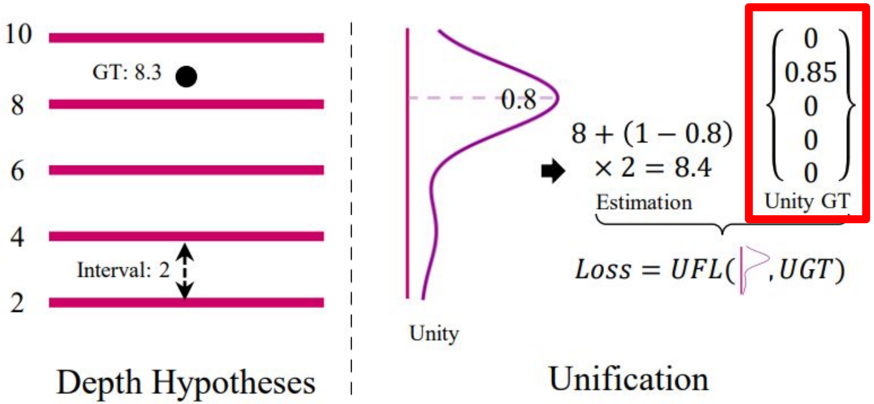


Method - Architecture

Cascade MVSnet + 2 additional modules + Unified Focal Loss



Method - Unity Generation



Algorithm 1: Unity Generation

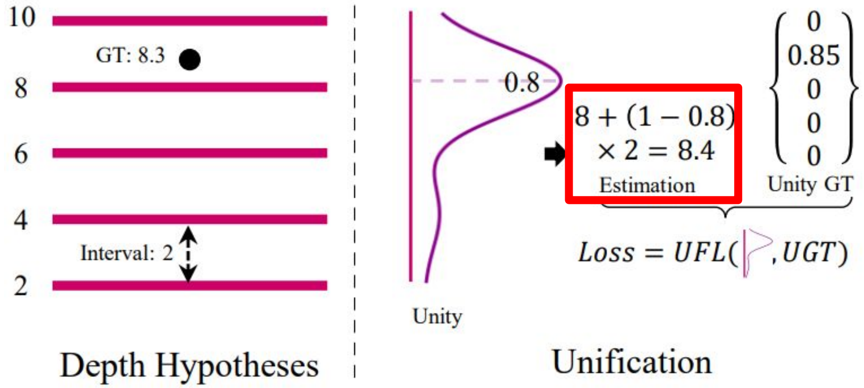
Input: Ground-truth depth $\mathbf{D}_{gt} \in \mathbb{R}^{H' \times W'}$; Depth hypotheses $\{\mathbf{d}_i \in \mathbb{R}^{H' \times W'}\}_{i=1}^{M}$.

Output: Ground-truth Unity $\{\mathbf{U}_i \in \mathbb{R}^{H' \times W'}\}_{i=1}^{M}$.

Initialization: Depth interval $r = 0$.

for $i = 1$ **to** M **do**
  for $(x, y) = (1, 1)$ **to** (H', W') **do**
    if $i < M$ **then**
      $r = \mathbf{d}_{i+1}^{x,y} - \mathbf{d}_i^{x,y}$;
    end
    if $\mathbf{d}_i^{x,y} \leq \mathbf{D}_{gt}^{x,y}$ **and** $\mathbf{d}_i^{x,y} + r > \mathbf{D}_{gt}^{x,y}$ **then**
      $\mathbf{U}_i^{x,y} = 1 - \frac{\mathbf{D}_{gt}^{x,y} - \mathbf{d}_i^{x,y}}{r}$;
    else
      $\mathbf{U}_i^{x,y} = 0$;
    end
  end
end
return $\{\mathbf{U}_i\}_{i=1}^{M}$.

Method - Unity Regression



Algorithm 2: Unity Regression

Input: Estimated unity $\{\widehat{\mathbf{U}}_i \in \mathbb{R}^{H' \times W'}\}_{i=1}^{M}$; Depth hypotheses $\{\mathbf{d}_i \in \mathbb{R}^{H' \times W'}\}_{i=1}^{M}$.
Output: Regressed Depth $\mathbf{D} \in \mathbb{R}^{H' \times W'}$.
Initialization: Depth interval $r = 0$.

```
for (x, y) = (1, 1) to (H', W') do
    Optimal hypothesis index o = argmax_{i∈{1,··· ,M}} Û_i^{x,y};
    Optimal hypothesis d = d_o^{x,y};
    if o < M then
        r = d_{o+1}^{x,y} − d_o^{x,y};
    else
        // previous interval for the last hypothesis
        r = d_o^{x,y} − d_{o−1}^{x,y};
    end
    Offset off = (1 − Û_o^{x,y}) × r;
    Depth D^{x,y} = d + off;
end
return D.
```

Method - Unified Focal Loss

Data Imbalance: α Hard Sample: γ

$$\mathrm{FL}(u, q) = \begin{cases} -\alpha(1-u)^{\gamma}\log(u), & q = 1 \\ -(1-\alpha)u^{\gamma}\log(1-u), & \text{else} \end{cases} \quad (6)$$

(Discrete Label)

Method - Unified Focal Loss

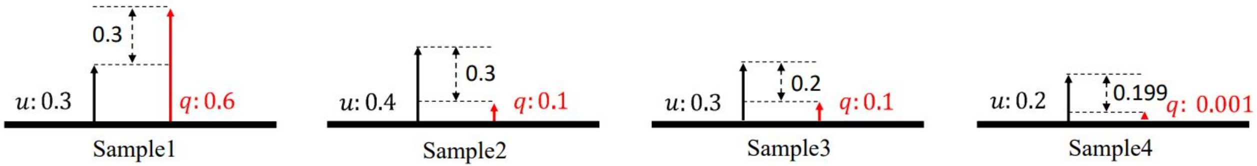
$$\mathrm{FL}(u,q) = \begin{cases} -\alpha(1-u)^{\gamma}\log(u), & q = 1 \\ -(1-\alpha)u^{\gamma}\log(1-u), & \text{else} \end{cases} \quad (6)$$

(Discrete Label)

$$\mathrm{GFL}(u,q) = \begin{cases} \alpha|q-u|^{\gamma}\mathrm{BCE}(u,q), & q > 0 \\ (1-\alpha)u^{\gamma}\mathrm{BCE}(u,q), & \text{else} \end{cases} \quad (7)$$

(Continuous Label)

Method - Unified Focal Loss



Absolute error : Sample1=Sample2>Sample3>Sample4	Relative error : Sample4>Sample2>Sample3>Sample1
0.3 0.3 0.2 0.199	199 3 2 0.5

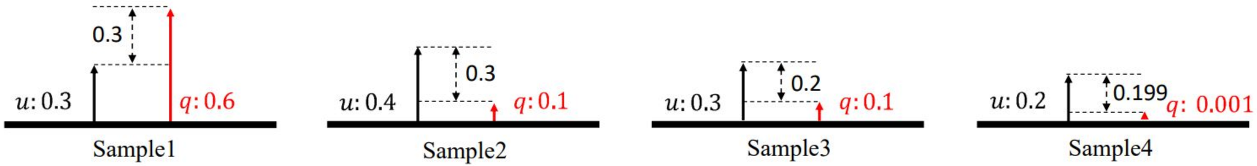
$$\mathrm{GFL}(u,q)=\begin{cases}\alpha|q-u|^{\gamma}\mathrm{BCE}(u,q), & q>0\\(1-\alpha)u^{\gamma}\mathrm{BCE}(u,q), & \text{else}\end{cases}\tag{7}$$

(Absolute error)

$$\mathrm{UFL}(u,q)=\begin{cases}\alpha(\frac{|q-u|}{q^{+}})^{\gamma}\mathrm{BCE}(u,q), & q>0\\(1-\alpha)(\frac{u}{q^{+}})^{\gamma}\mathrm{BCE}(u,q), & \text{else}\end{cases}\tag{8}$$

(Relative error)

Method - Unified Focal Loss



Absolute error : Sample1=Sample2>Sample3>Sample4 0.3 0.3 0.2 0.199	Relative error : Sample4>Sample2>Sample3>Sample1 199 3 2 0.5

$$\mathrm{UFL}(u,q)=\begin{cases}\alpha(\frac{|q-u|}{q^+})^\gamma \mathrm{BCE}(u,q), & q>0\\ (1-\alpha)(\frac{u}{q^+})^\gamma \mathrm{BCE}(u,q), & \text{else}\end{cases} \quad (8)$$

(Relative error)

$$\mathrm{UFL}(u,q)=\begin{cases}\alpha^+(S_b^+(\frac{|q-u|}{q^+}))^\gamma \mathrm{BCE}(u,q), & q>0\\ \alpha^-(S_b^-(\frac{u}{q^+}))^\gamma \mathrm{BCE}(u,q), & \text{else}\end{cases} \quad (9)$$

(Relative error with sigmoid)

Our Proposed Improvements

Optimization of the Calculation of Unification & Unified Focal Loss

$$U_i^{x,y} = 1 - \frac{\mathrm{D}_{gt}^{x,y} - d_i^{x,y}}{r} \quad off = \left(1 - \widehat{U}_o^{x,y}\right) * r$$

$$U_i^{x,y} = 1 - \frac{\mathrm{D}_{gt}^{x,y} - d_i^{x,y}}{2 * r} \quad off = \left(1 - \widehat{U}_o^{x,y}\right) * (2 * r)$$

Our Proposed Improvements

Optimization of the Calculation of Unification & Unified Focal Loss

$$\mathrm{UFL}(u,q)=\begin{cases}\alpha^{+}(S_b^{+}(\frac{|q-u|}{q^{+}}))^{\gamma}\mathrm{BCE}(u,q),\\ \alpha^{-}(S_b^{-}(\frac{u}{q^{+}}))^{\gamma}\mathrm{BCE}(u,q),\end{cases}$$

$$S_5^{+}(x)=4\times(\frac{1}{1+5^{-x}}-0.5)+1 \quad [1,3)$$

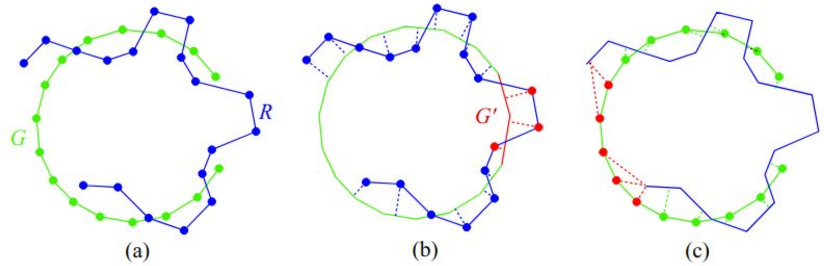
$$S_5^{-}(x)=2\times(\frac{1}{1+5^{-x}}-0.5) \quad [0,1)$$

q: unity ground truth
u: estimation depth

Expected Result

- Improve the problem of inaccurate prediction in the early training period.
- Conduct correction on balancing the positive and hard samples without sigmoid functions
- Obtain better results by trying out different activation functions

Metrics



- **Accuracy** : distance from the MVS reconstruction to the structured light reference, encapsulating the quality of the reconstructed MVS points.

- **Completeness** : distance from the reference to the MVS reconstruction, encapsulating how much of the surface is captured by the MVS reconstruction.

img reference: S. M. Seitz, B. Curless, J. Diebel, D. Scharstein and R. Szeliski, "A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms," *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, 2006, pp. 519-528

Result

	Baseline	Baseline + OUG	Opti. Baseline	Opti. Baseline w/o Sigmoid	Opti. Baseline w/o Sigmoid range [1, 3]	Baseline w/ ReLU	Baseline w/ Leaky ReLU
Accuracy	1.6048	0.4623	0.419	0.3886	0.4179	0.3915	0.4175
Completeness	1.1208	0.3322	0.3283	0.3229	0.3472	0.4722	0.5392
Overall	1.3628	0.3973	0.3737	0.3557	0.3826	0.4318	0.4783

Conclusion

- Outperform the baseline (original UniMVSNet) , improve overall performance by 73.89%.

- Optimized Baseline converges faster, with training time per epoch reduced from 5 hours to 4 hours (on RTX 2080 Ti).

- The overall performance improved by 68% after replacing the loss function in the baseline model.

Thank you for your listening

Any Questions?