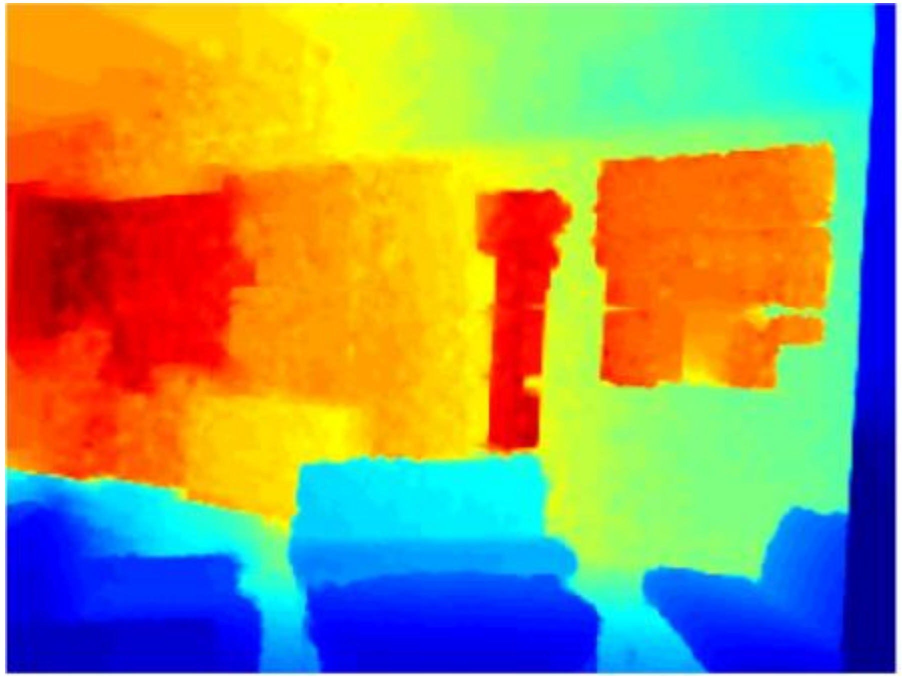
Monocular Depth Estimation Based on Multi-Scale Graph Convolution Networks

Wooseok Oh

Seoul National University



Motivation to use GCN

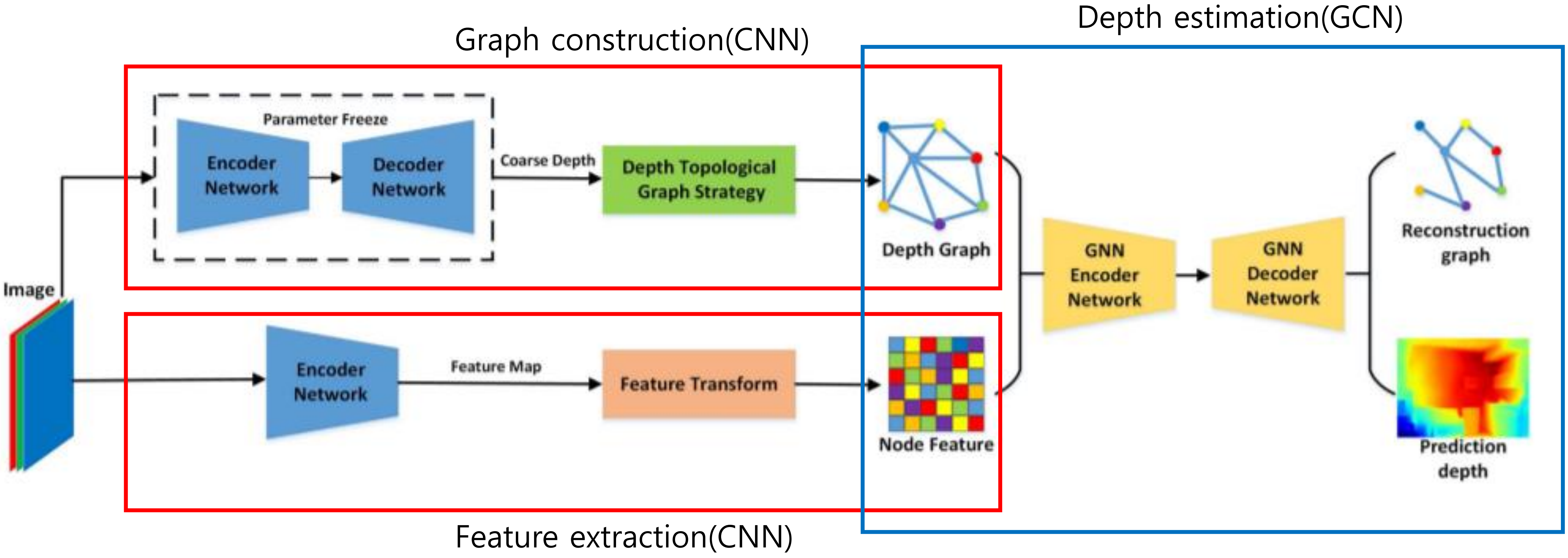
- CNN can only extract the feature of receptive field of filter
- Depth value is related to
  - The neighboring pixel values → CNN
  - Other pixels of the same depth value → GCN

Neighbor & not related

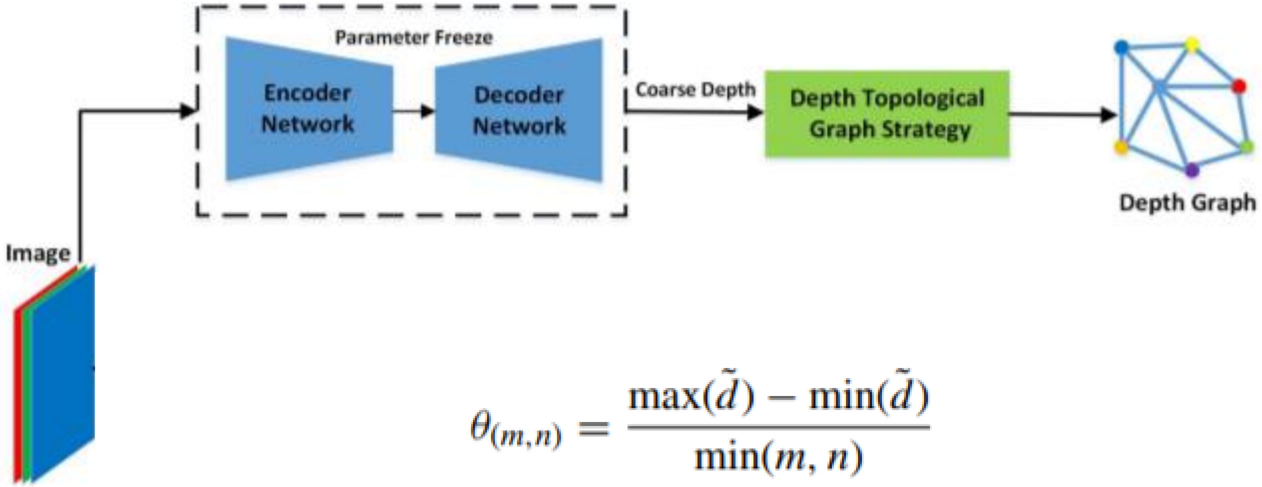
Not neighbor & related

Neighbor & related

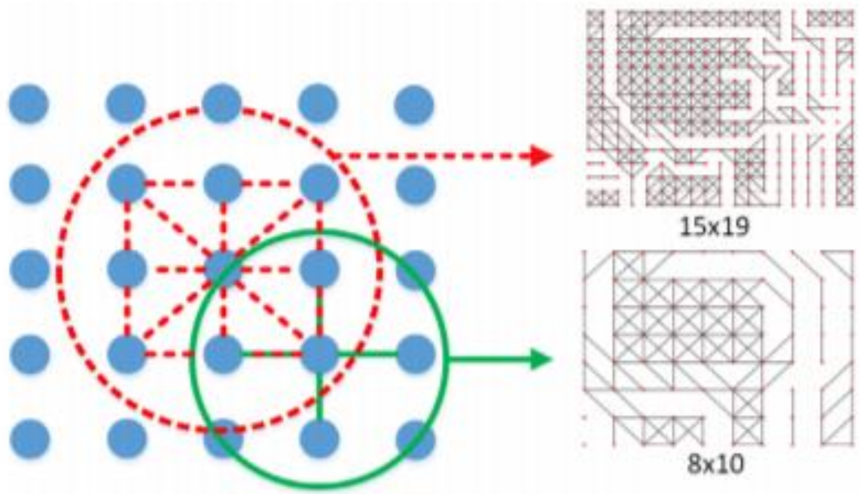
Network Architecture



Graph Construction



$$\theta_{(m,n)} = \frac{\max(\tilde{d}) - \min(\tilde{d})}{\min(m, n)}$$



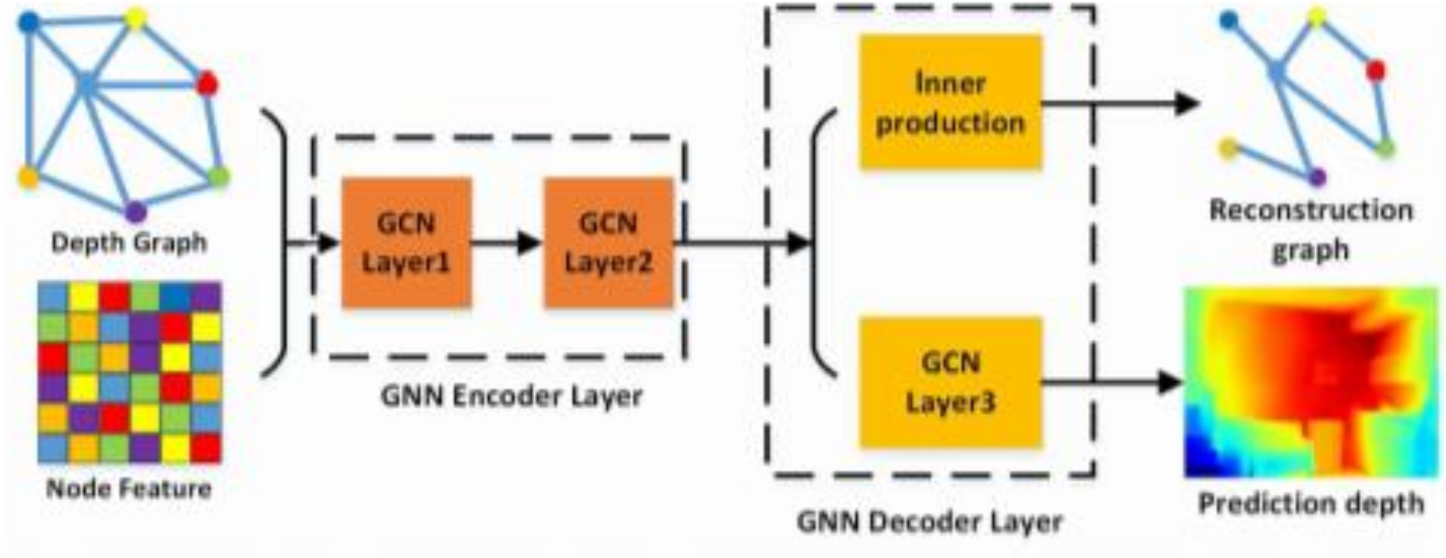
Graph construction using threshold θ
where $\tilde{d}$ is the coarse-grained depth map. m is the rows number. n is the columns number.

Feature extraction

Use same encoder in coarse depth prediction network

Extract multi-scale node feature in encoder

GCN for depth estimation



$$\tilde{A} = A + I \quad \textbf{(1)}$$

$$H^{l+1} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{l}W^{l}\right) \quad \textbf{(2)}$$

$$\hat{A} = sigmoid(H^{l+1}H^{l+1^{T}}) \quad \textbf{(3)}$$

A : adjacency matrix of graph

$\widetilde{D}$: degree matrix of $\tilde{A}$ (all weights are same)

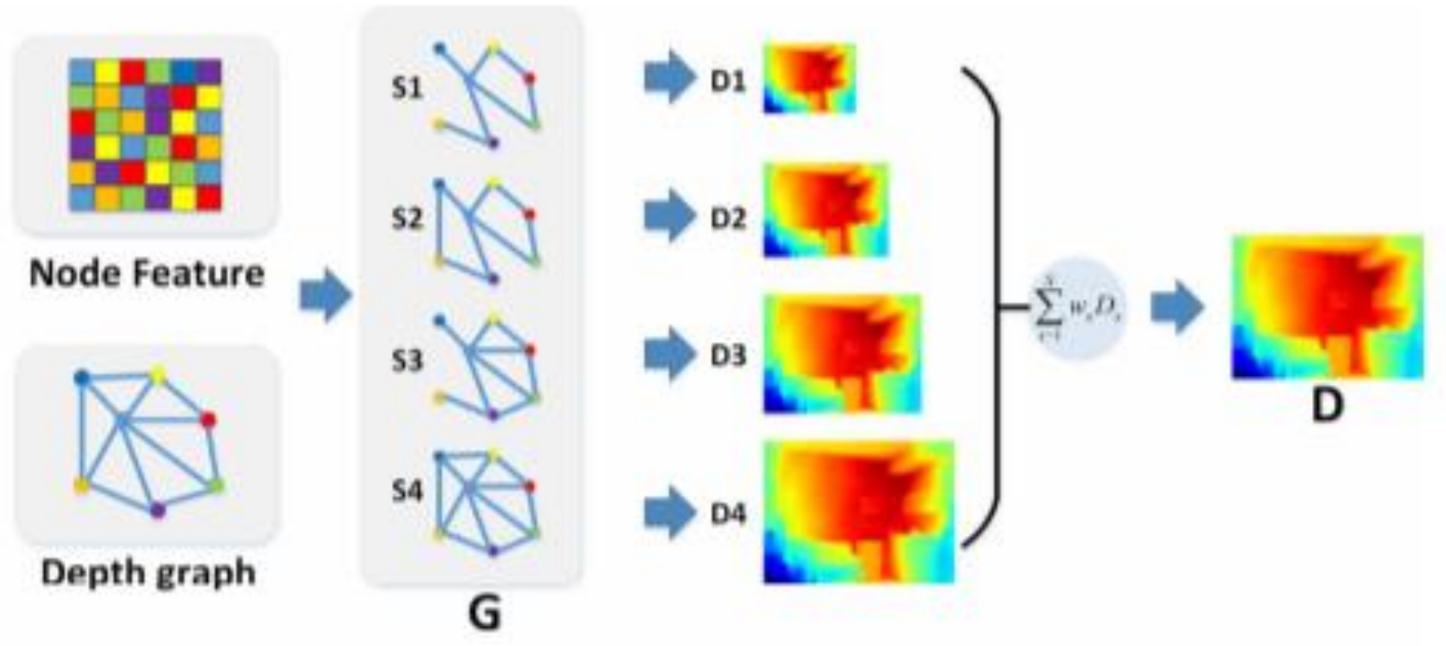
H^l: l'th layer's node feature. (N x D matrix)

(H^0 = feature extracted by CNN)

W^l: trainable weight matrix

$\hat{A}$: reconstruction graph's adjacency matrix

Multiscale depth estimation



Loss function

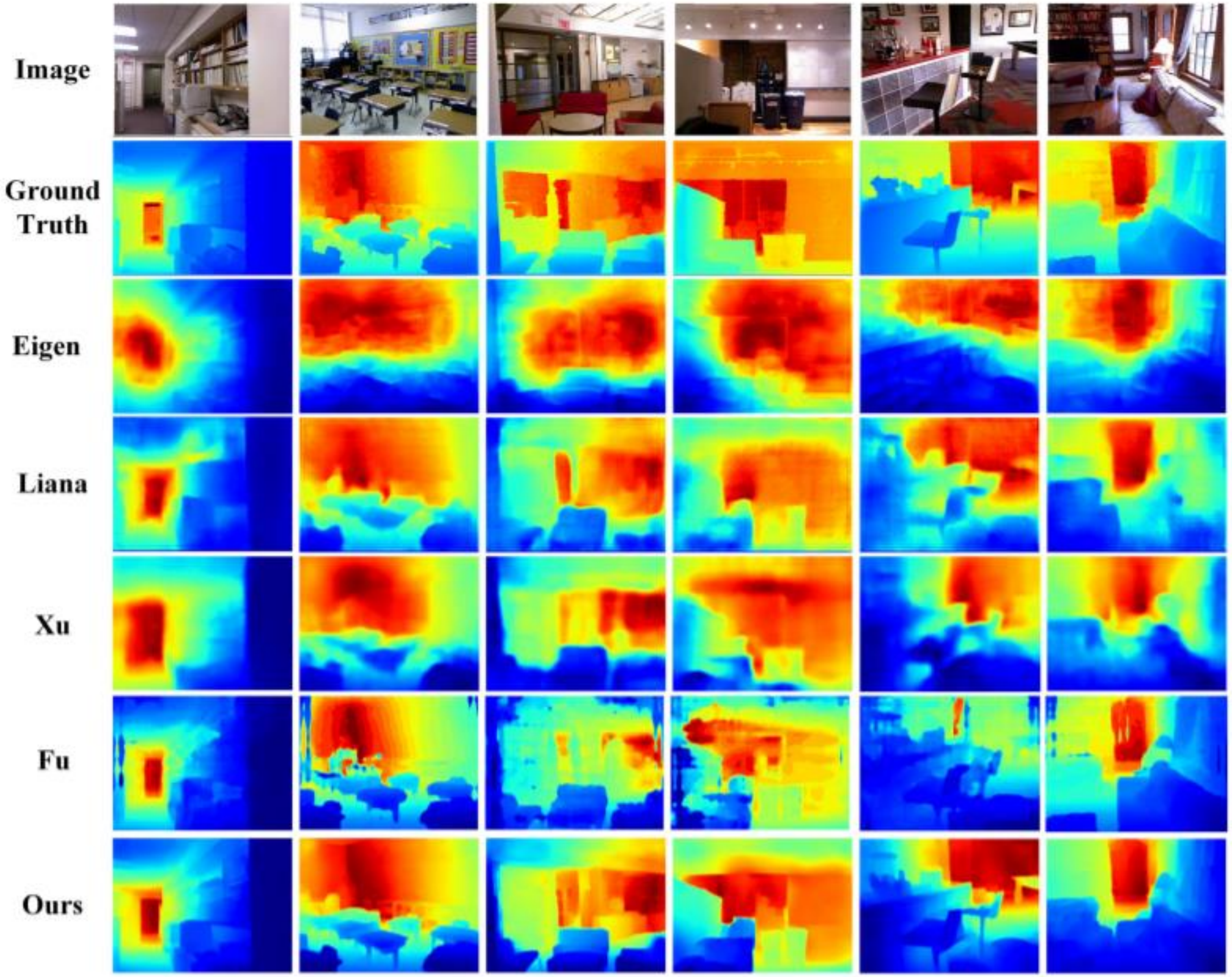
$$l_{dice} = 1 - \frac{1}{S}\sum_{s=1}^{S} Dice_s \quad Dice_s = \frac{2|\hat{A}_s \bigcap A_s|}{|\hat{A}_s| + |A_s|}$$

$$l_{grad} = \frac{1}{n}\sum_{i=1}^{n}\left[(\nabla_x d_i)^2 + (\nabla_y d_i)^2\right]$$

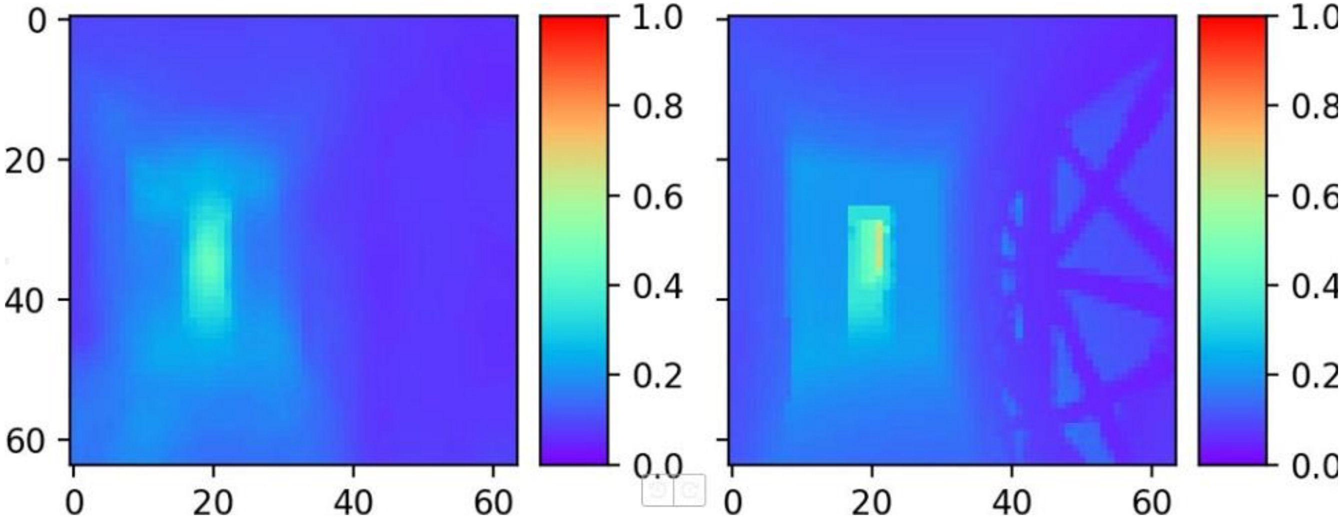
$$l_{si} = \frac{1}{n}\sum_i d_i^2 - \frac{1}{n^2}\left(\sum_i d_i\right)^2$$

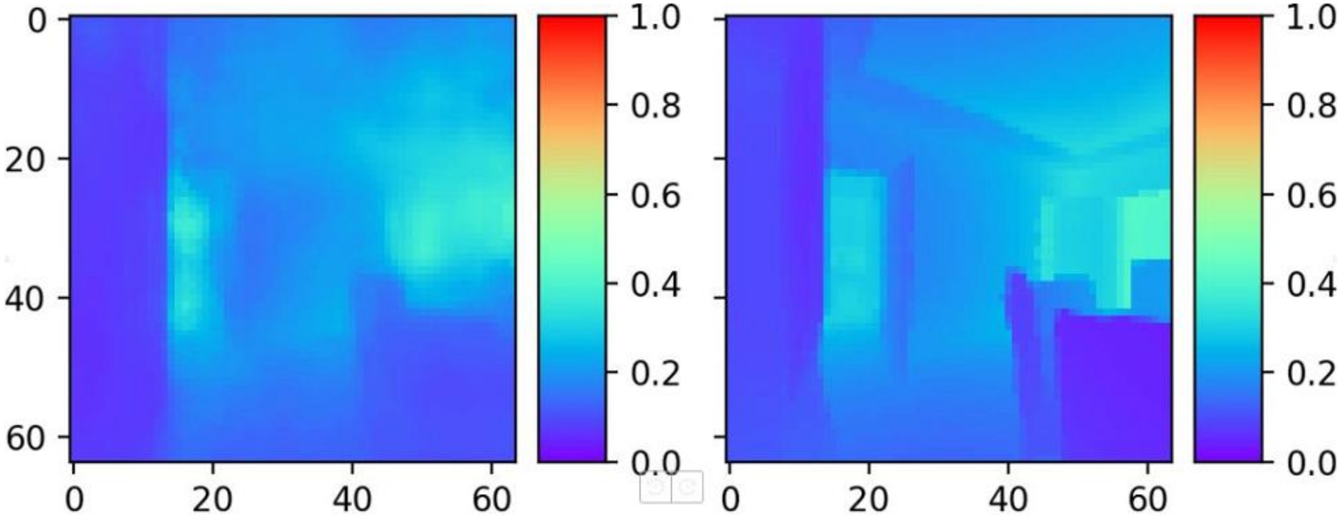
$$L = l_{si} + l_{grad} + \lambda l_{dice}$$

Result



Implementation result





Result of implementation

- 코드는 직접 구현해서, 발표날에 이미 구현 결과를 ppt에 올려서 발표했다.
- 성능이 생각보다 좋지 않게 나와서 몇번 더 실험 해 봤지만, 더 좋은 결과를 얻지는 못했다.
- 앞 페이지에 있는 결과가 가장 잘 학습된 네트워크에서 나온 결과이다.
- 여러 번 실험하면서 같은 network parametr 파일에 overwrite 해서 가장 좋은 성능을 보인 네트워크의 파라미터는 남아있지 않다.
- 학습 과정에 저장된 rgb 이미지와 depth 이미지는 압축해서 제출했으니, 앞쪽의 이미지 외의 결과가 더 궁금하다면 그 이미지를 확인해 보길 바람/