Neural Image Representations for Multi-Image Fusion and Layer Separation

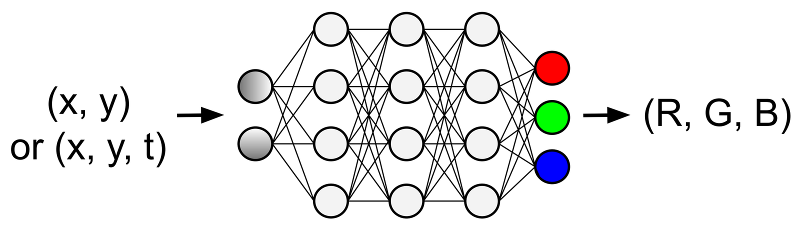
Seonghyeon Nam Marcus A. Brubaker Michael S. Brown
York University

https://shnnam.github.io/research/nir

TEL AVIV 2022

Motivation

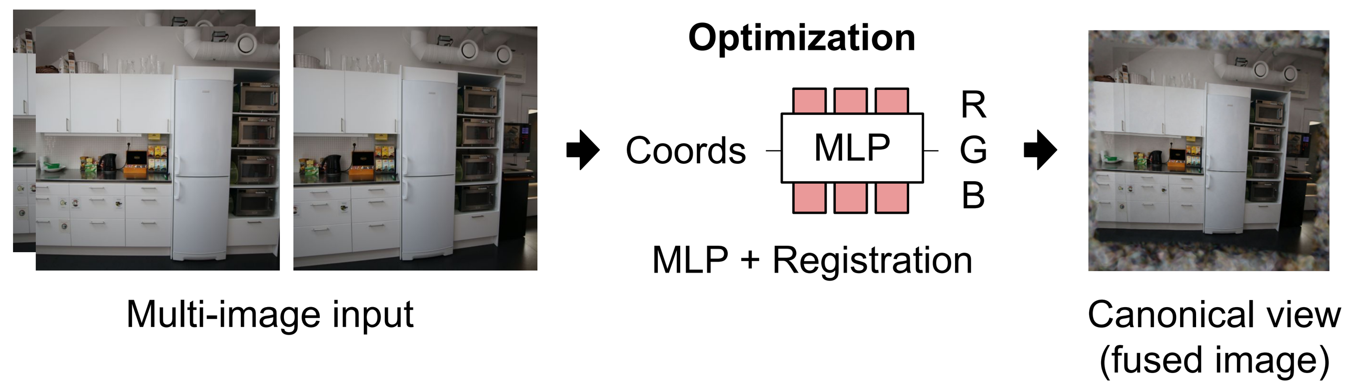
❑ Implicit (or coordinate-based) neural representations



Multi-layer perceptron (MLP)

- Image signal is continuous with respect to pixel coordinates
  - Can access to real-valued pixel positions
  - Can interpolate image signal at such locations
- Several works studied image processing applications [1, 2, 3]

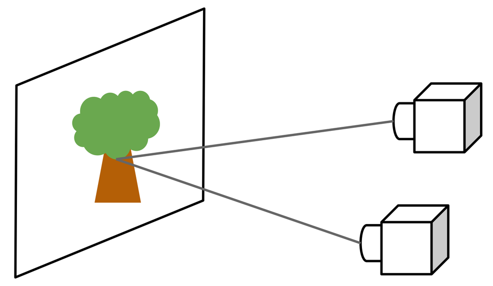
❑ Neural image representations (NIRs)



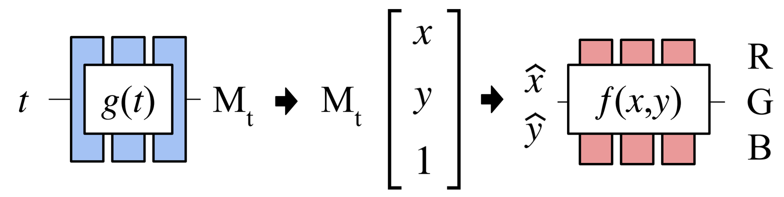
- Multi-image fusion in the continuous space represented by an MLP
  - No explicit reference image and interpolation
- Application to multi-image layer separation in an unsupervised manner
  - e.g. burst demoireing, obstruction removal, deraining

Neural Image Representations (NIRs) for Multi-Image Fusion

❑ Homography-based NIRs

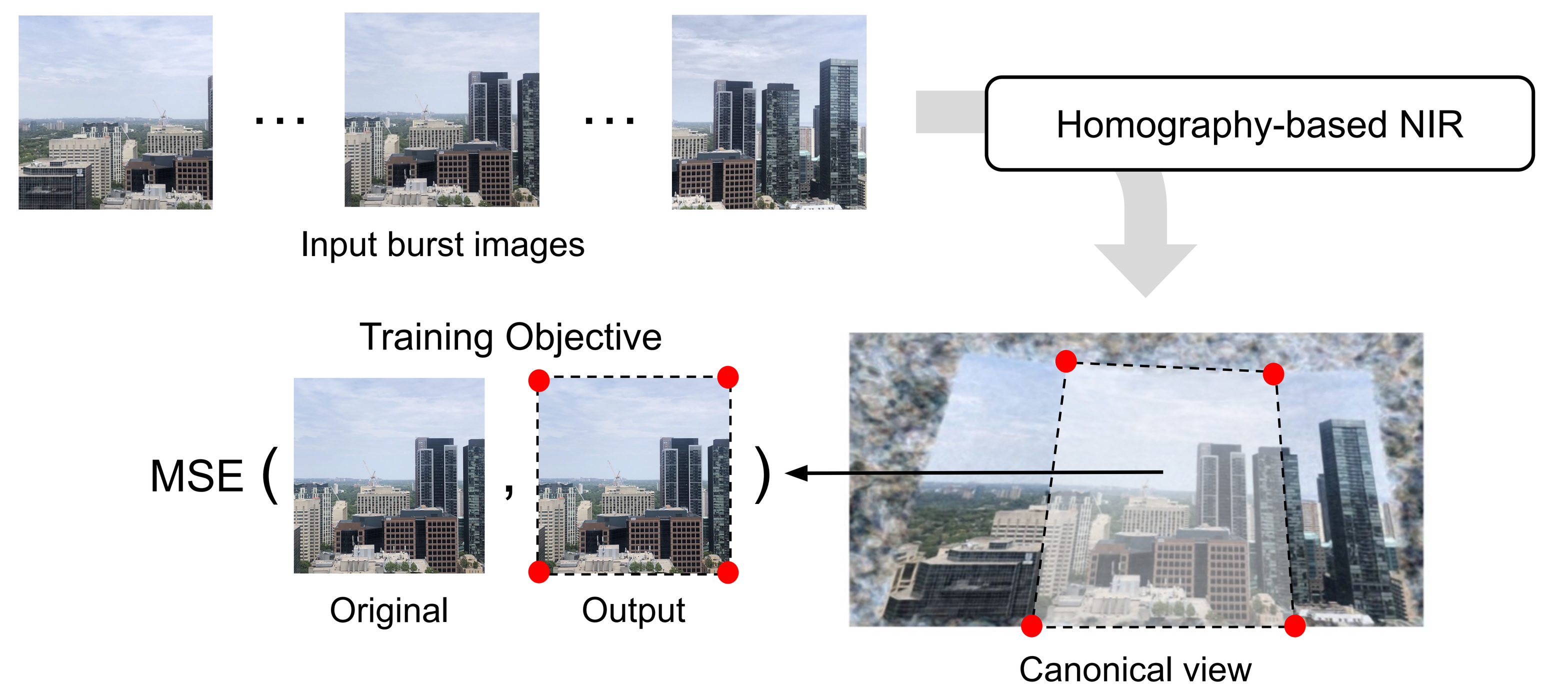


Planar, rigid scenes

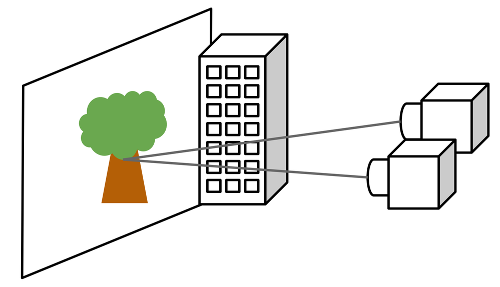


- Planar motion, no occlusion
- An MLP $f(x,y)$ learns a canonical view where all frames are fused
- Another MLP $g(t)$ estimates a 3x3 homography matrix, which transforms pixel coordinates at t to the canonical view
- Both MLPs are learned by a self-reconstruction loss:

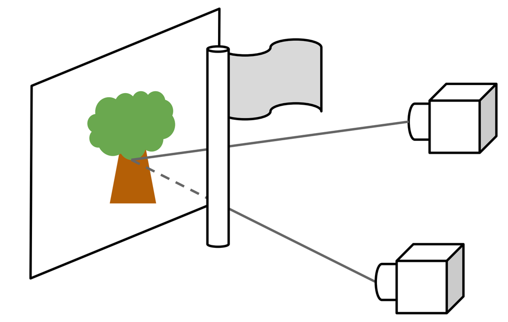
$$\mathcal{L}_{\text{Recon}} = \sum_{x,y,t} \|\hat{\mathbf{I}}_{(x,y,t)} - \mathbf{I}_{(x,y,t)}\|_2^2$$



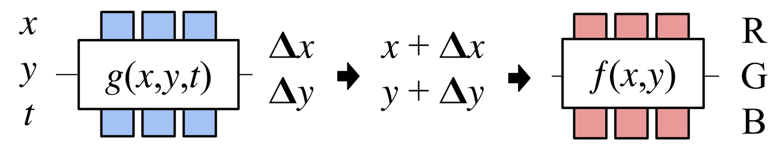
❑ Flow-based NIRs



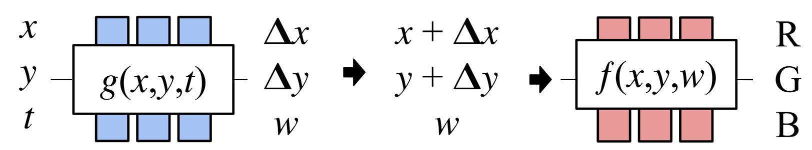
Non-planar, rigid scenes with small motion



Non-planar, non-rigid scenes with large motion



Regularization: $\sum \|J_g(x, y, t)\|_1$



Regularization: $\sum \|J_g(x, y, t)\|_1$

- **Flow-based occlusion-free NIRs**
- Non-planar motion, no occlusion
- Regularization for a flow map

- **Flow-based occlusion-aware NIRs**
- Non-planar motion, occlusion
- w stores occlusion/disocclusion

Input

xy-flow map (left) and *w* map (right)

Reconstruction

Canonical views at t = 0, 2, 3

Application to Multi-Image Layer Separation

❑ Two-stream NIRs

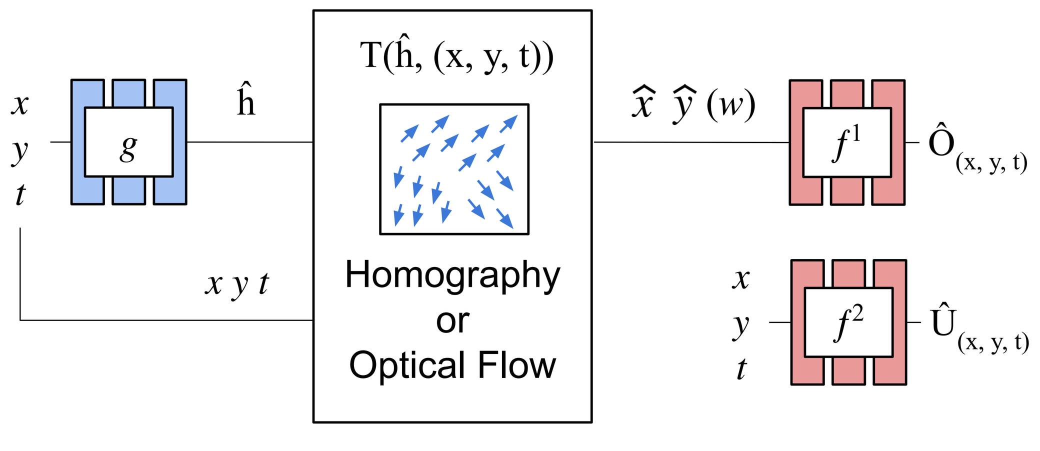


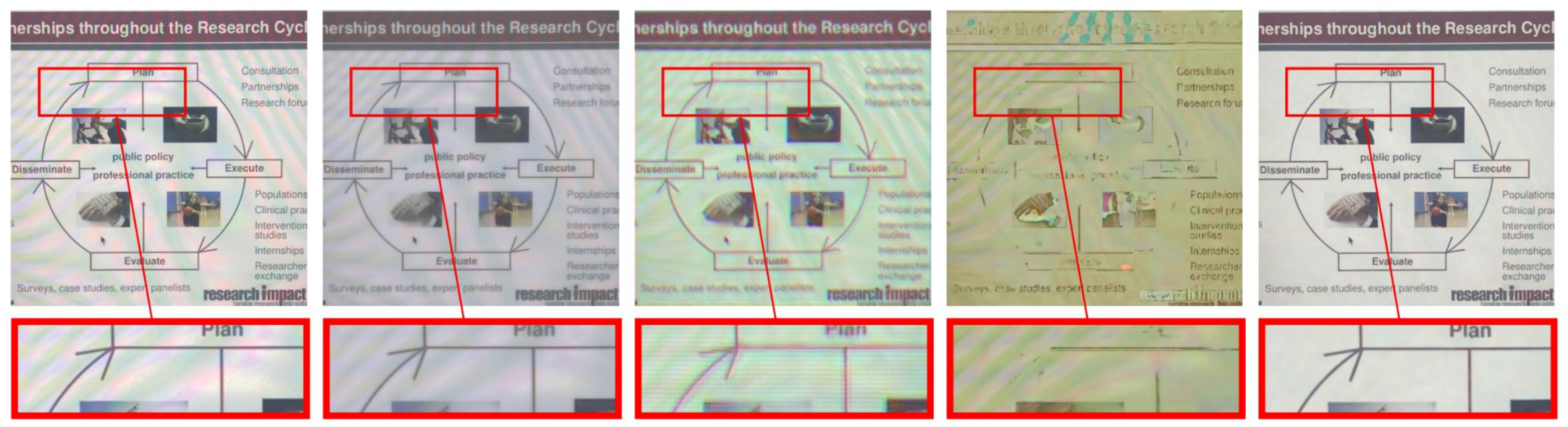
Image formation: $\hat{\mathrm{I}}_{(x,y,t)} = \hat{\mathrm{O}}_{(x,y,t)} + \hat{\mathrm{U}}_{(x,y,t)}$

- Separate image signals into two layers in f^1 and f^2
- Assume that two layers have different motion
- One layer that follows the motion $g(x,y,t)$ is learned in f^1, while the other layer is stored in f^2
- For each layer separation task, task-specific regularization losses are added
  (Please refer to the paper for the details)

❑ Applications

All qualitative results are obtained using 5 consecutive real images as input

- **Moire removal**



Input AFN C3Net Double DIP Ours

- **Reflection removal**

Input Li and Brown Alayrac *et al.* Double DIP Liu *et al.* Ours

- **Fence removal**

Input Liu *et al.* Liu *et al.* Ours

- **Rain removal**

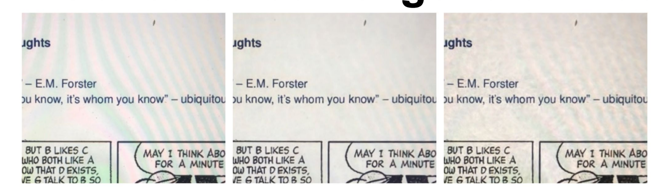
Input FastDeRain SpacCNN Ours

❑ Discussion

- **Different number of images**

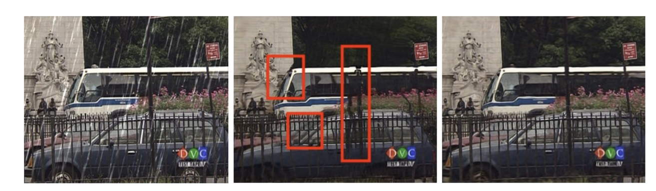
Input 2 images 5 images

- **Constrained vs. general model**



Input Homography Flow

- **Effect of *w***



Input w/o *w* w/ *w*

References
[1] Chen et al. Learning Continuous Image Representation with Local Implicit Image Function. In CVPR, 2021.
[2] Chen et al. VideoINR: Learning Video Implicit Neural Representation for Continuous Space-Time Super-Resolution. In CVPR, 2022.
[3] Mai et al. Motion-Adjustable Neural Implicit Video Representation. In CVPR, 2022.