JUNE 2022

IMPLICIT NEURAL REPRESENTATION AND UNIFIED IMPLICIT NEURAL STYLIZATION

PRESENTER: ZHIWEN FAN
PhD Student, The University of Texas at Austin
Why Go Implicit?

Image(2D, discrete)  
Mesh(3D, discrete)  
Audio(discrete)

If images weren’t discrete grids, would we use CNN today?
How to go infinite resolution for representing signals?
Why Go Implicit?

Image(2D, discrete)

Mesh(3D, discrete)

Audio(discrete)
Why Go Implicit?

Input (360px)  Pixels  Bilinear Interpolation  Implicit-based [1]

[1] Chen et al., Learning Continuous Image Representation with Local Implicit Image Function
Benefits (e.g., Signed Distance Function-SDF)

Benefits:
- Agnostic to the resolution
- Model memory scales with signal complexity
Categories

Image Fitting (SIREN)
(Sitzmann et al, 2020 [1])
(x, y) → pixel intensity

Deep SDF
(Park et al, 2019 [2])
(x, y, z) → distance

Neural Radiance Field
(Ben et al, 2020 [3])
(x, y, z) → (color, density)

Scene Representation Network
(Sitzmann et al, 2019 [3])
(x, y, z) → latent vector (color, dist)

[1] Sitzmann et al., Implicit Neural Representations with Periodic Activation Functions
[2] Park et al., Deepsdf: Learning continuous signed distance functions for shape representation
[3] Ben et al., Representing Scenes as Neural Radiance Fields for View Synthesis
INR (Image Fitting)

[1] Sitzmann et al., Implicit Neural Representations with Periodic Activation Functions
INR (Signed Distance Function/SDF)

\[ \hat{x} = c + t_0 v - \frac{v}{\nabla f_0 \cdot v_0} f(c + t_0 v) \]
\[ \hat{n} = \nabla f(\hat{x}) \]

\[(x, y, z) \rightarrow \text{distance}\]

[1] Yariv et al, Multiview Neural Surface Reconstruction by Disentangling Geometry and Appearance
INR (Neural Radiance Fields/NeRF)

Input Images $\rightarrow$ Optimize NeRF $\rightarrow$ Render new views

Neural Radiance Field
INR (Neural Radiance Fields/NeRF)

$F_\theta \rightarrow (r, g, b, \sigma)$

- $(x, y, z, \theta, \phi)$
  - Spatial location
  - Viewing direction
- $F_\theta$:
  - Fully-connected neural network
  - 9 layers,
  - 256 channels
- Output color
- Output density
NeRF Preliminary: Volume Rendering

Rendering model for ray $r(t) = o + td$:

$$C \approx \sum_{i=1}^{N} T_i \alpha_i c_i$$

- $C$: colors
- $T_i$: weights
- $N$: number of segments

How much light is blocked earlier along ray:

$$T_i = \prod_{j=1}^{i-1} (1 - \alpha_j)$$

accumulated transmittance along the ray

How much light is contributed by ray segment $i$:

$$\alpha_i = 1 - e^{-\sigma_i \delta t_i}$$
NeRF: Training Pipeline
Question: How to Edit INR?
Question: How to Edit INR?
Unified Implicit Neural Stylization (INS)

Image Fitting

Signed Distance Function

Neural Radiance Field
INS Framework

- Style Implicit Module
  - \((s_0, s_1, \ldots, s_N)\) Style Embedding
  - MLP
- Content Implicit Module
  - \((x, y, z)\) Coordinates
  - MLP

- Amalgamate Module
- MLP
- Image

- SDF
- SDF Rendering
- NeRF
- Volumetric Rendering

- Direct
- Losses:
  - \(L_{\text{style}}\)
  - \(L_{\text{content}}\)
INS Framework

Inputs: implicit coordinates, ray directions and style embeddings. Style Implicit Module (SIM) and Content Implicit Module (CIM) are used to extract conditional implicit style features and implicit scene features.
Amalgamate Module (AM) is applied to fuse features, generating stylized density and color intensity of each 3D point.
An implicit rendering step is applied on the top of AM to render the pixel intensity.
INS Results (Image Fitting)

Content+Style Image

Stylized Image

Stylized Image
INS Results (SDF)

Content Image (1 view)  Stylize Image  Stylized Image
The proposed self-distilled geometry consistency. The CIM weights in the grey box are from a pre-trained NeRF and are kept fixed since then. During fine-tuning, the output density of the fixed CIM serves as geometry constrains for the stylized density from the output of AM.
INS Framework (NeRF)

Qualitative results of the self-distilled geometry consistency ("GC"). Both the rendered images and depth maps are shown to validate the effectiveness of the proposed consistency.
INS Framework (NeRF)

Sampling Stride = 1  
Sampling Stride = 2  
Sampling Stride = 4

Color Image  Style  Color w/ SS  Color w/o SS  Depth w/ SS  Depth w/o SS
INS Results (NeRF)

Content Image

Stylize Image

Stylization
INS Results (NeRF)

Content Image

Stylize Image

Stylization
INS Results (NeRF)

Content Image

Stylize Image

Stylization
INS Results (NeRF Style Interpolation)
INS Results (NeRF Comparisons)

Content Scene and Style Image

Perceptual
[Xun et.al, 2017]

MCCNet
[Deng et.al, 2021]

Ours

Style3D
[Chiang et.al, 2021]

ReReVST
[Wang et.al, 2020]
INS Results (NeRF Comparisons)

Content Scene and Style Image

Ours

Perceptual
[Xun et.al, 2017]

MCCNet
[Deng et.al, 2021]

Style3D
[Chiang et.al, 2021]

ReReVST
[Wang et.al, 2020]
INS Results (NeRF Comparisons)

Content Scene and Style Image

Ours

Style3D
[Chiang et.al, 2021]

Perceptual
[Xun et.al, 2017]

MCCNet
[Deng et.al, 2021]

ReReVST
[Wang et.al, 2020]
Conclusion

- Implicit Neural Representation (INR), a better continuous representation for small scale object/scene.
- We can edit INR on the feature domain to preserve its continuity.

- Implicit Neural Representation (INR) typically requires per-scene training, which is not very generalizable.
What’s Next

- NeRF for dynamic scenes.
- NeRF for better view-dependent effect & light transport.
- INR with mid/high level tasks.
Thank you!

Zhiwen Fan (zhiwenfan@utexas.edu, https://zhiwenfan.github.io/)