



SCHOOL OF ELECTRICAL &
COMPUTER ENGINEERING
Aristotle University of Thessaloniki

3D Ανακατασκευή Επιφανειών μέσω Έμμεσων Αναπαραστάσεων

Έμφαση στην κωδικοποίηση υψηλοσυχνοτικού
περιεχομένου

Χάρης Φίλης ΑΕΜ: 9449,

Επιβλέποντες:

Αναστάσιος Ντελόπουλος, Καθηγητής
Αντώνης Καρακώπτας, Υποψήφιος Διδάκτωρ





Εισαγωγή

3D Ανακατασκευή
Αναπαραστάσεις Γραφικών

Μεθοδολογία

Neural Implicit Rendering
IDR
Fourier Neural Tangent
Kernel
Multi-Resolution 3D Hash
Encoding
Neural Fourier Filter Banks

Πειράματα, Αποτελέσματα, Αξιολόγηση

Dataset
Learning Curves
Αποτελέσματα Rendering
Μετρικές Αξιολόγησης
Φωτογραφίας
Μετρικές Αξιολόγησης
Γεωμετρίας

Συμβολή & Μελλοντικές Επεκτάσεις

Συμβολή
Προβλήματα
Μελοντικές Επεκτάσεις

Εισαγωγή

Διατύπωση προβλήματος 3D ανακατασκευής



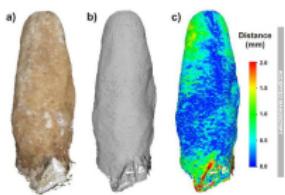
2

- Ανακατασκευή 3D μοντέλων από 2D εικόνες

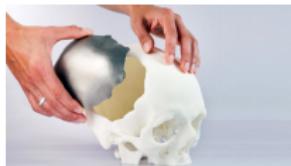


Είσοδος συλλογή 2D εικόνων

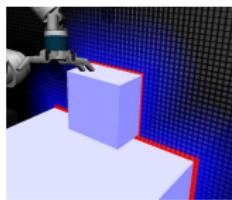
Έξοδος 3D Μοντέλο



Γεωδαισία



Ιατρική
3D
εκτυπωμένα
εμφυτεύματα,
Εκμαγεία



Ρομποτική
3D Αντίληψη
Χώρου



Τέχνη, Ιστορία

<https://doi.org/10.1016/j.cageo.2011.09.012>

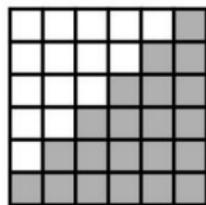
<https://arxiv.org/abs/2204.02296>

<https://doi.org/10.1016/j.media.2008.12.003>

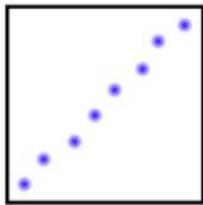
<https://doi.org/10.1111/cgf.12077>



Voxel



Points



Meshes



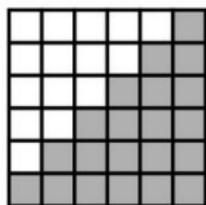
Εισαγωγή

Αναπαραστάσεις Γραφικών

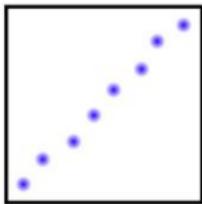


4

Voxel



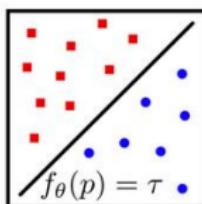
Points



Meshes



Implicits



Εισαγωγή

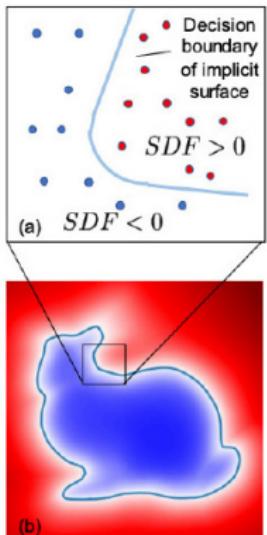
Γιατί χρησιμοποιούνται Έμμεσες Αναπαραστάσεις;



- Μπορούν να αναπαρασταθούν(U.A.T.) ως ένα σύνολο εξόδου νευρωνικού δικτύου (Διαφορίσιμη Πολλαπλότητα - Neural Level Set)
- Αναπαριστούν τον χώρο μέσω SDF(**Signed Distance Function**)
- Η επιθυμητή επιφάνεια ανιστοιχεί στην Ισομετρική Επιφάνεια (Zero Level Set):

$$S_\theta = \{\mathbf{x} \in \mathbb{R}^3 | f(\mathbf{x}; \theta) = 0\}$$

- Συντεταγμένες $(x, y, z) \in \mathbb{R}^3 \rightarrow \mathbb{R}$ (SDF Value)
- Δεν διακριτοποιούν το πρόβλημα της ανακατασκευής
- Κωδικοποιούν την εμφάνιση του αντικειμένου



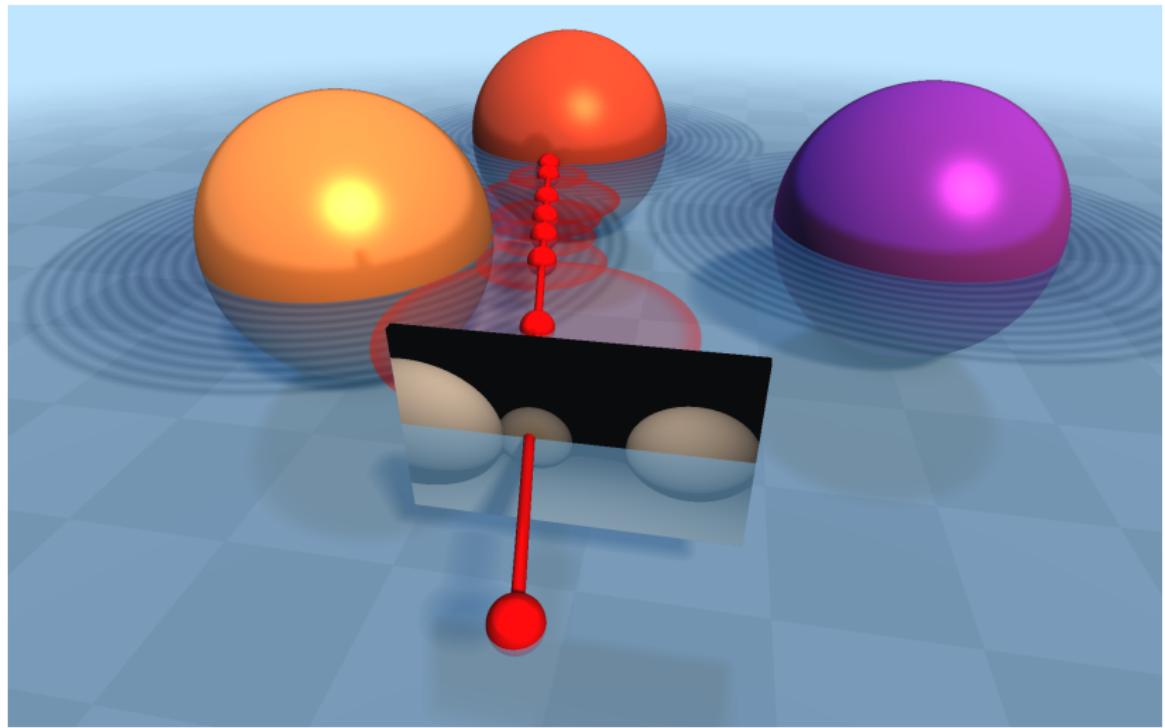
SDF

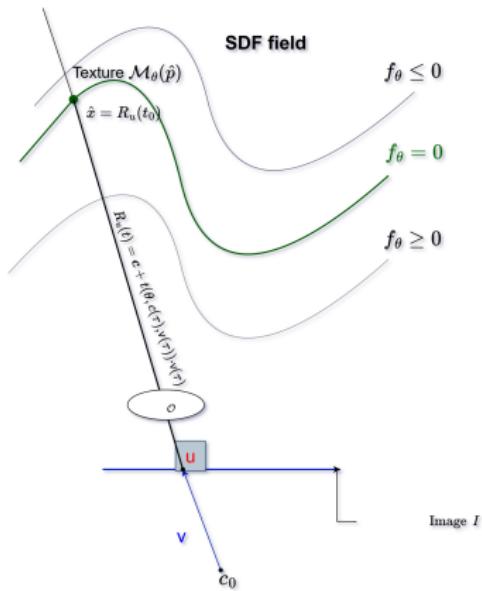
Εισαγωγή

Οπτική Απόδοση Αλγορίθμου Sphere Tracing σε πεδίο SDF



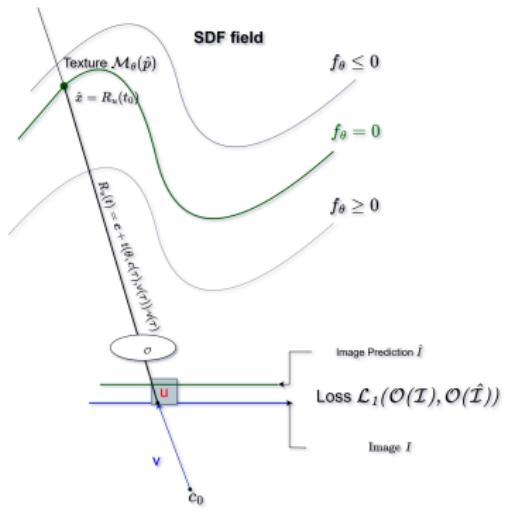
6





Forward Pass:

- $\forall u \in O(I)$
- Sphere Tracing για εύρεση \hat{x} πάνω στην ακτίνα $R_u(t) = c + t \cdot v$
- Αν $f_\theta \leq 0$ εντός , $f_\theta \geq 0$ εκτός της επιφάνειας
- Για \hat{x} (σημείο τομής) με εκτιμώμενη προσημασμένη απόσταση $f(\hat{x})$
- Χρώμα Texture Field $\mathcal{M}_\theta(\hat{x})$ στο σημείο \hat{x}
- Εκτίμηση χρώματος pixel $O(\hat{I}_u)$ με αντίστροφή αποτύπωση



Backward Pass:

- $\mathcal{L}(\mathcal{O}(I), \mathcal{O}(\hat{I})) = \sum_{O_u} |\hat{I}_u - \hat{I}|$
 $\forall u \in O(I)$
- Επιστροφή κλίσης σφαλμάτων στα βάρη των \mathcal{M}, f δικτύων
- Κανόνας Αλυσίδας

$$\frac{\partial \mathcal{L}}{\partial \theta} = \sum_{\mathcal{O}_u} \frac{\partial \mathcal{L}}{\partial \hat{\mathbf{I}}_u} \cdot \frac{\partial \hat{\mathbf{I}}_u}{\partial \theta}$$

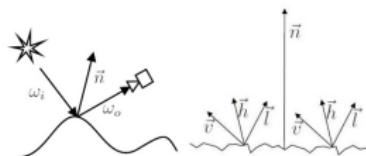
$$\frac{\partial \hat{\mathbf{I}}_u}{\partial \theta} = \frac{\partial \mathbf{M}_\theta(\hat{x})}{\partial \theta} + \frac{\partial \mathbf{M}_\theta(\hat{x})}{\partial \hat{x}} \cdot \frac{\partial \hat{x}}{\partial \theta}$$
- Έμμεση Διαφόριση σημείων τομής \hat{x} της εξίσωσης
 $f_\theta(\hat{x}(\theta, \tau); \theta) = 0,$
 $f_\theta(R_u(c + t(\theta, c(\tau), v(\tau)) * v(\tau)); \theta) = 0$



Πεδίο Ακτινοβολίας Επιφάνειας

$$L(\hat{x}, w^o) = L^e(\hat{x}, w^o) + \int_{\Omega} B(\hat{x}, \hat{n}, w^i, w^o) L^i(\hat{x}, w^i)(\hat{n}, w^i) dw^i$$

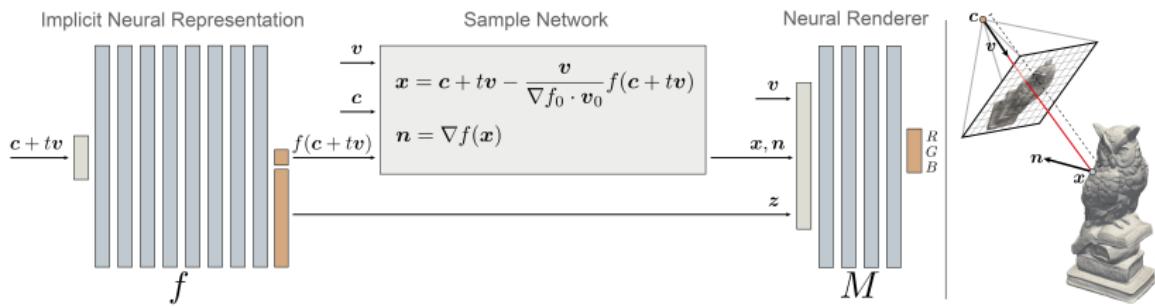
- $L(\hat{x}, w^o)$: Εκπέμπουσα ακτινοβολία από την σκηνή
- $B(\hat{x}, \hat{n}, w^i, w^o)$: Συνάρτηση κατανομής ανάκλασης διπλής κατεύθυνσης
- $L^i(\hat{x}, w^i)$ Προσλαμβανόμενη ακτινοβολία στην επιφάνεια από άλλες πηγές φωτός
- \hat{n}, w^i : κανονικά διανύσματα και διανύσματα εκπομπής (αποσβεστικοί παράγοντες στο μοντέλο κατοπτρικής ανάκλασης)
- Ω : Σφαιρικό χώριο επικεντρωμένο στο κανονικό διάνυσμα \hat{n}



Restricted Lightfield - P Universality

Το συνεχές μέρος των P συναρτήσεων M_0 περιγράφεται από ένα επαρκώς βαθύ MLP δίκτυο

$$L(\theta, \gamma, \tau) = M(\hat{x}, \hat{n}, v; \gamma)$$





Συνολική Συνάρτηση Κόστους

$$\text{loss}(\theta, \gamma, \tau) = \text{loss}_{\text{RGB}}(\theta, \gamma, \tau) + \rho \text{ loss}_{\text{MASK}}(\theta, \tau) + \lambda \text{ loss}_{\text{E}}(\theta)$$



Φωτομετρικό Σφάλμα RGB (Εμφάνιση)

$$\text{Loss}_{\text{RGB}}(\theta, \gamma, \tau) = \frac{1}{|P|} \sum_{u \in P^{\text{in}}} |I_u - L_u(\theta, \gamma, \tau)|$$

Cross Entropy - Σφάλμα Μάσκας (Γεωμετρία)

$$\text{Loss}_{\text{MASK}}(\theta, \tau) = \frac{1}{\alpha |P|} \sum_{u \in P^{\text{out}}} \text{CE}(\mathcal{O}_u, \mathcal{S}_{u,\alpha}(\theta, \tau))$$

Παράγοντας Κανονικοποίησης Σημείων - Σφάλμα Εικονικής Εξίσωσης (SDF)

$$\text{Loss}_{\text{Eikonal}}(\theta) = \mathbb{E}_x \left(\|\nabla_x f(x; \theta)\| - 1 \right)^2$$



Προβλήματα Coordinate based MLP

- Τα δίκτυα συντεταγμένων $F(x, y, z; \theta)$ είναι **MLP** δίκτυα με **ReLU, SoftPlus** συναρτήσεις ενεργοποίησης ή γενικότερα τμηματικά γραμμικές συναρτήσεις που δεν έχουν την ικανότητα (**capacity**) να αποτυπώσουν υψηλοσυχνοτικές λεπτομέρειες της εικόνας στην έμμεσα αναπαριστώμενη επιφάνεια.
- Δεν λαμβάνονται υπόψιν σχέσεις που υπάρχουν σε υψηλότερες διαστάσεις μεταξύ των δεδομένων.



Προβλήματα Coordinate based MLP

- Τα δίκτυα συντεταγμένων $F(x, y, z; \theta)$ είναι **MLP** δίκτυα με **ReLU, SoftPlus** συναρτήσεις ενεργοποίησης ή γενικότερα τμηματικά γραμμικές συναρτήσεις που δεν έχουν την ικανότητα (**capacity**) να αποτυπώσουν υψηλοσυχνοτικές λεπτομέρειες της εικόνας στην έμμεσα αναπαριστώμενη επιφάνεια.
- Δεν λαμβάνονται υπόψιν σχέσεις που υπάρχουν σε υψηλότερες διαστάσεις μεταξύ των δεδομένων.

Εκπαίδευση

Το πρόβλημα έχει τις ρίζες του και στην ίδια την δομή των νευρωνικών δικτύων και την διαδικασία της εκπαίδευσης



Προβλήματα Coordinate based MLP

- Τα δίκτυα συντεταγμένων $F(x, y, z; \theta)$ είναι **MLP** δίκτυα με **ReLU, SoftPlus** συναρτήσεις ενεργοποίησης ή γενικότερα τμηματικά γραμμικές συναρτήσεις που δεν έχουν την ικανότητα (**capacity**) να αποτυπώσουν υψηλοσυχνοτικές λεπτομέρειες της εικόνας στην έμμεσα αναπαριστώμενη επιφάνεια.
- Δεν λαμβάνονται υπόψιν σχέσεις που υπάρχουν σε υψηλότερες διαστάσεις μεταξύ των δεδομένων.

Εκπαίδευση

Το πρόβλημα έχει τις ρίζες του και στην ίδια την δομή των νευρωνικών δικτύων και την διαδικασία της εκπαίδευσης

Προτεινόμενη λύση

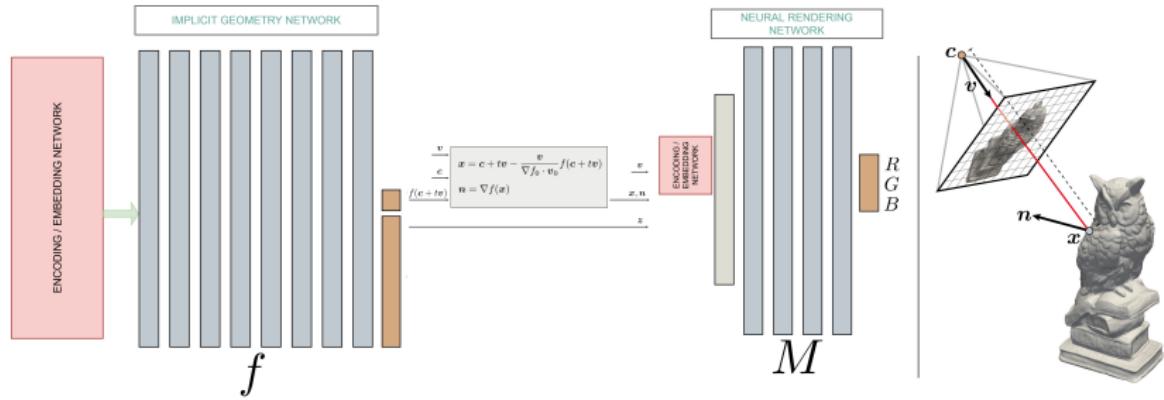
Κωδικοποίηση-Μετασχηματισμός Εισόδου(**Input Embedding**)

Χρήση Υψηλοσυχνοτικής Κωδικοποίησης

Neural Frequency Input Embedding

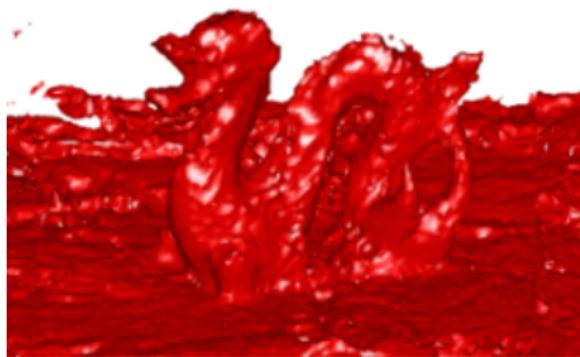


14





Truncated SDF



Χρήση Περικοπής SDF $\tanh(b \frac{x}{2})$, με x τιμή SDF, b Laplace Density του $L_D(x)$



Πυρήνας Ημιτονοειδών Συναρτήσεων Fourier

- $\vec{\gamma}(v) = [\alpha_1 \cdot \cos(2\pi b_1^T v), \alpha_2 \cdot \sin(2\pi b_2^T v)], \dots, \alpha_m \cdot \cos(2\pi b_m^T v)]^T$
- Ιδιότητα **Στασιμότητας**

$$K_\gamma = \gamma(v_1)^T \cdot \gamma(v_2) = \sum_{j=1}^m a_j^2 \cos(2\pi b_j^T (v_1 - v_2)) = h_\gamma(v_1 - v_2)$$



Fourier Features Variations

- Απλές Ημιτονοειδές Συναρτήσεις

$$\gamma(\mathbf{v}) = [\cos(2\pi\mathbf{v}), \sin(2\pi\mathbf{v})]^T$$

- Positional Encoding - Λογαριθμική Δειγματοληψία Συχνοτήτων οι οποίες εφαρμόζονται με βάση την θέση

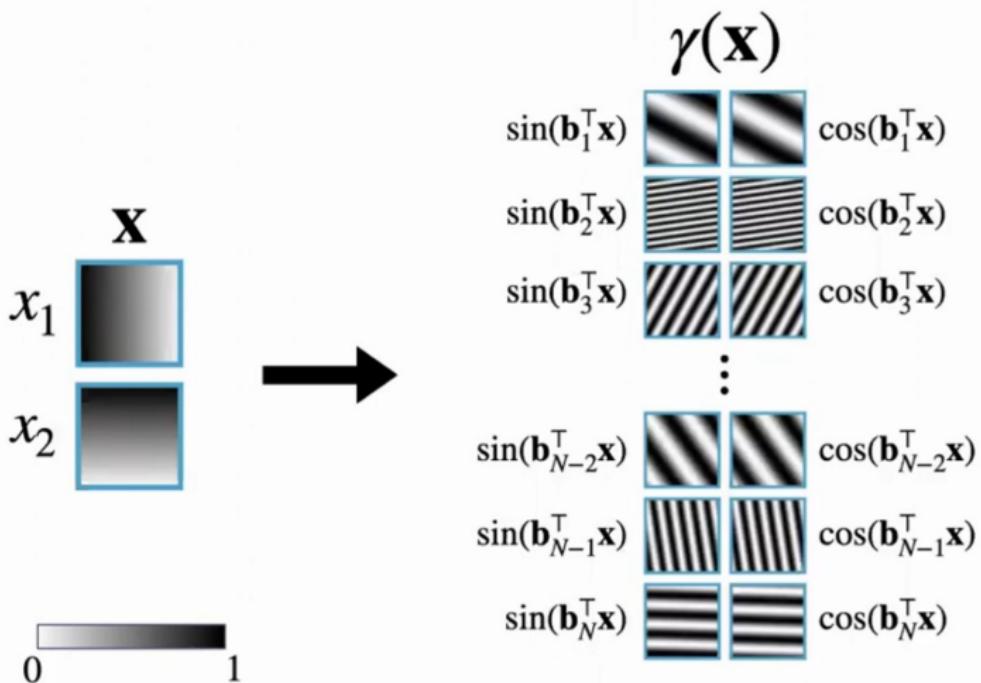
$$\gamma(\mathbf{v}) = [\dots, \cos(2\pi\sigma^{j/m}\mathbf{v}), \sin(2\pi\sigma^{j/m}\mathbf{v}), \dots]^T \text{ για } j = 0, \dots, m-1$$

- Fourier Features - Δειγματοληψία συχνοτήτων από Τυχαία Γκαουσιανή Κατανομή \mathbf{B}

$$\gamma(\mathbf{v}) = [\cos(2\pi\mathbf{B}\mathbf{v}), \sin(2\pi\mathbf{B}\mathbf{v})]^T$$

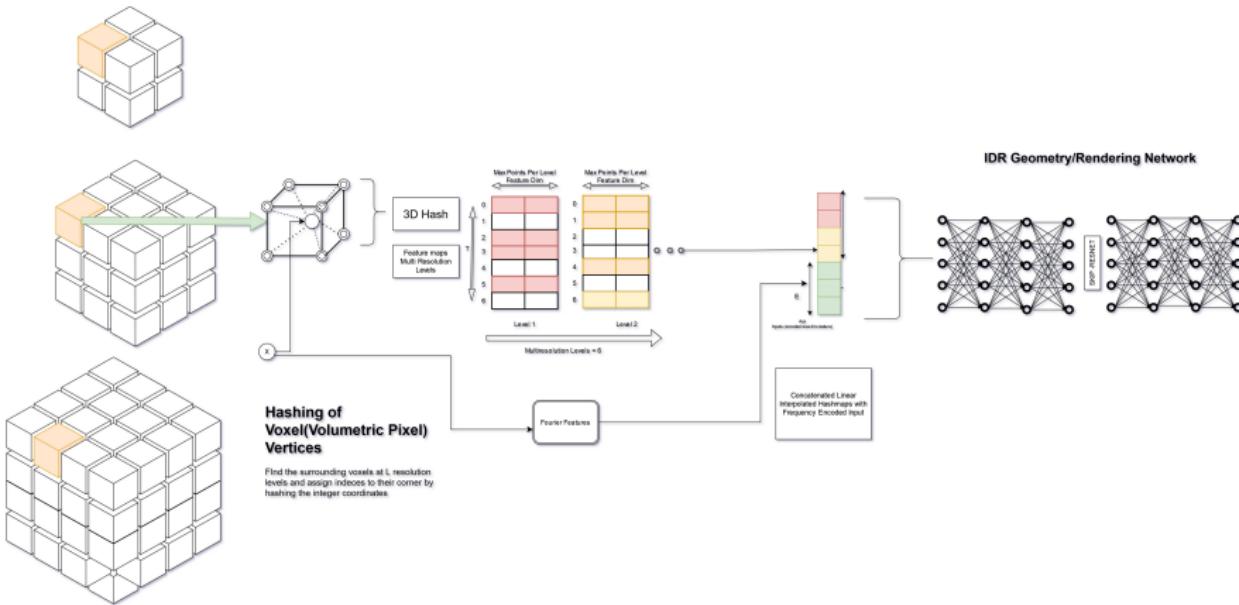


Fourier feature mapping: simple 2D example



Multi-Resolution Hash Encoding

Χωρικός Κατακερματισμός Πολλαπλών Επιπέδων Ανάλυσης

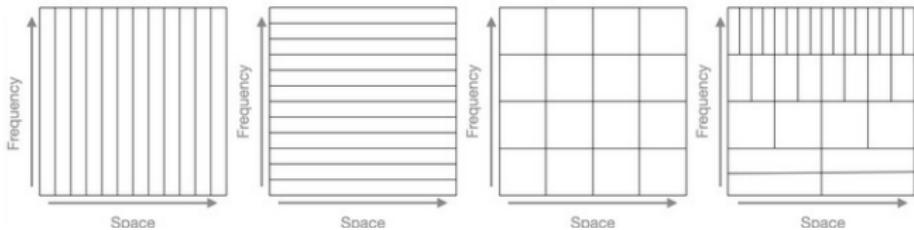
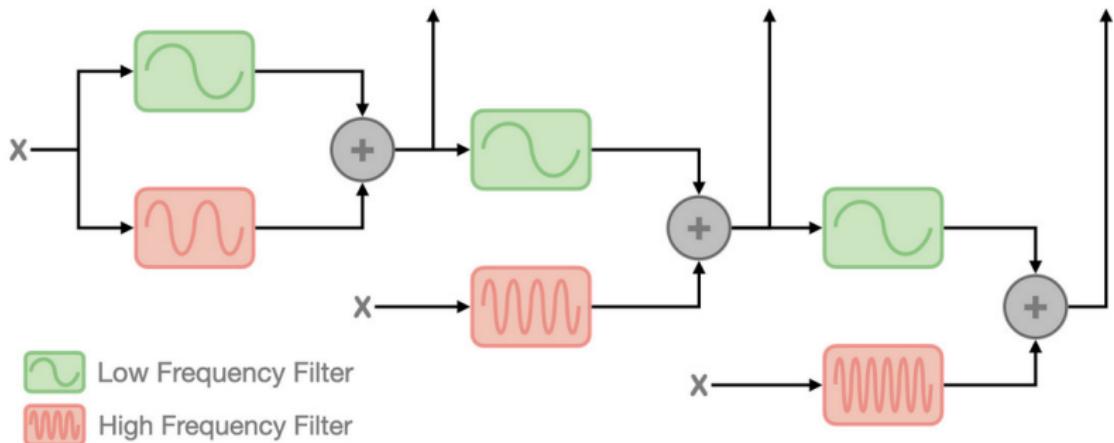


Μεθοδολογία

Έμπνευση: Μετασχηματισμός Κυματιδίων



20

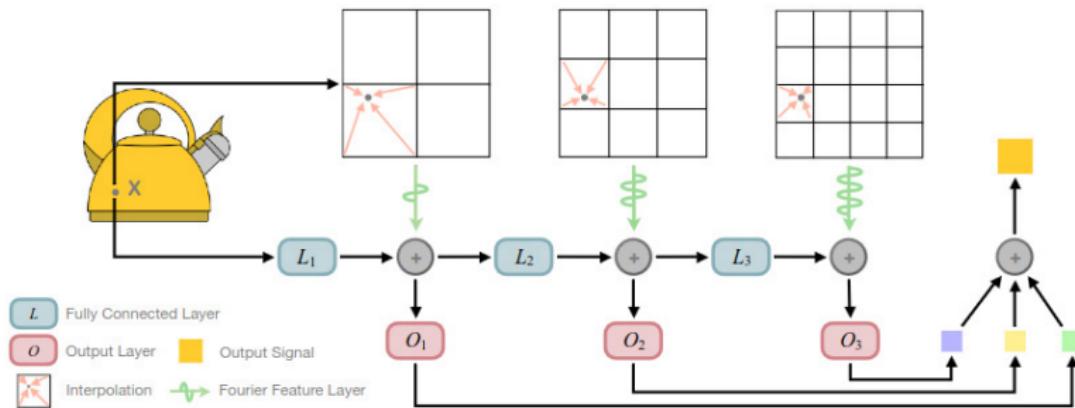


Neural Fourier Filter Banks

Βαθιά Νευρωνική Κωδικοποίηση Θυρίδων Fourier



21

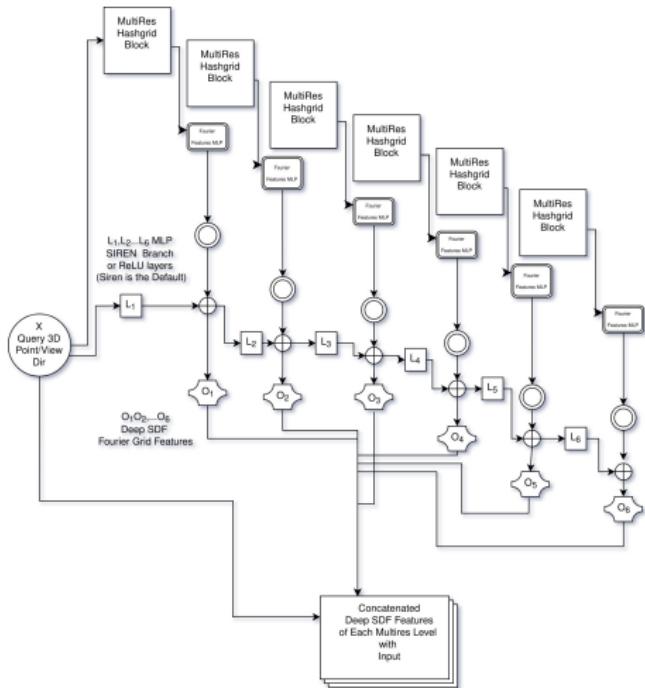


Neural Fourier Filter Banks

Βαθιά Νευρωνική Κωδικοποίηση Θυρίδων Fourier



22

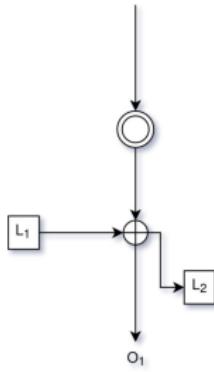


Style Weight Modulation/Demodulation

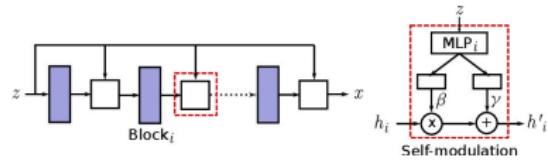
Διαμόρφωση βαρών με βάση Style Χαρακτηριστικών



23



Style mod/demod block



Διαμόρφωση Βαρών



Σύνολο δεδομένων DTU, 64/49 Θέσεων Κάμερας



Σκηνές Εκπαίδευσης



Σκηνή 65
Κρανίο



Σκηνή 110
Χρυσός Λαγός



Σκηνή 114
Κεραμικός
Βούδας



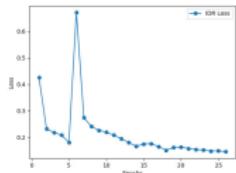
Σκηνή 122
Πήλινη
Κουκουβάγια

Πειράματα

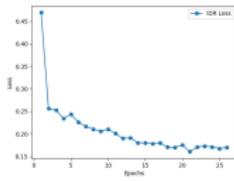
[Positional Encoding](#) | [Fourier Features](#) | [Multi-Resolution 3D HashGrid Encoding](#) | [NFFB](#) | [StylemodNFFB](#) | [HashGrid3D-TinyCUDANN](#) | [StylemodNFFB-TinyCUDANN](#)

Καμπύλες Εκπαίδευσης

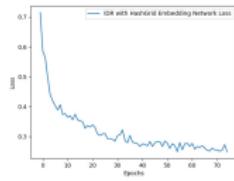
Σύγκριση Σύγκλισης Αλγορίθμων [Σκηνή 65 - Κρανίο]



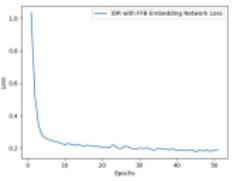
Positional
Encoding



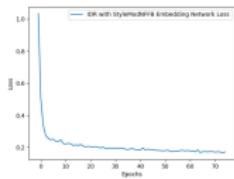
Fourier
Features



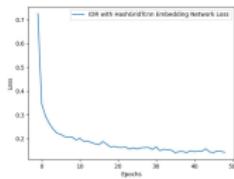
MR 3D
HashGrid



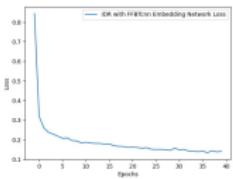
NFFB



StylmodNFFB



MR 3D Hashgrid
TCNN



StylemodNFFB TCNN

Καμπύλες Εκπαίδευσης

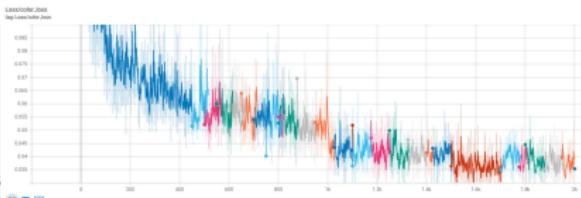
Καμπύλες 2000 Εποχών Εκπαίδευσης [Δίκτυο NFFB - Σκηνή 65 - Κρανίο]



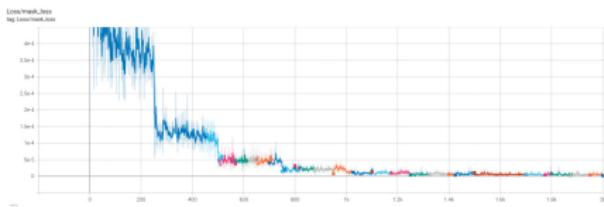
27



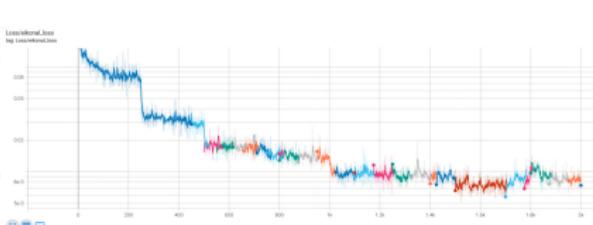
Συνολικό Σφάλμα



RGB Σφάλμα



Σφάλμα Μάσκας Ενεργών
Pixel Σκηνής



Σφάλμα Εικονικής Εξίσωσης

Αποτελέσματα Αποτύπωσης

Rendering Evaluation Results - Σκηνή 122 Κουκουβάγια



28



Positional Encoding



Fourier Features



MR 3D HashGrid



NFFB



StylmodNFFB



MR 3D Hashgrid TCNN



StylemodNFFBTCNN

Αποτελέσματα Αποτύπωσης

Rendering Evaluation Results - Σκηνή 122 Κουκουβάγια - Zoomed



29



Positional Encoding



Fourier Features



MR 3D HashGrid



NFFB



StylmodNFFB



MR 3D Hashgrid TCNN



StylemodNFFBTCNN

Πίνακες Φωτομετρικής Αξιολόγησης

PSNR - Λόγος σήματος κορυφής προς θόρυβο



30

Embedding Algorithm	PSNR Score : Owl[122]
Positional Encoding	27.15
Fourier Features	26.51
Multi-Resolution Hash Grid	25.85
Multi-Resolution Hash Grid (<i>TinyCudaNN</i>)	27.26
NFFB	27.4
StyleMod NFFB	28.04
StyleModNFFB (<i>TinyCudaNN</i>)	27.57

Αποτελέσματα αξιολόγησης με τη μετρική PSNR (DTU Scene: Owl[122])

Πίνακες Φωτομετρικής Αξιολόγησης

SSIM - Structure Similarity Index Measure



31

Embedding Algorithm	SSIM Score : Skull[65]
Positional Encoding	0.95
Fourier Features	0.94
Multi-Resolution Hash Grid	0.95
NFFB	0.94
StyleMod NFFB	0.94
StyleModNFFB (<i>TinyCudaNN</i>)	0.96

Αποτελέσματα αξιολόγησης με τη μετρική SSIM(DTU Scene: Skull[65])

Πίνακες Φωτομετρικής Αξιολόγηση

LPIPS - Learned Perceptual Image Patch Similarity (Alexnet)



32

Embedding Algorithm	LPIPS Score: Rabbit[110]	LPIPS Score: Buda[114]
Positional Encoding	0.15	0.19
Fourier Features	0.16	0.10
Multi-Resolution Hash Grid	0.14	0.19
Multi-Resolution Hash Grid (<i>TinyCudaNN</i>)	0.11	
NFFB	0.03	0.22
StyleMod NFFB	0.16	0.20
StyleModNFFB (<i>TinyCudaNN</i>)	0.15	0.16

Αποτελέσματα αξιολόγησης με τη μετρική LPIPS (DTU Scenes: Rabbit[110] and Buda[114])

Chamfer Distance

Αξιολόγηση Γεωμετρίας Ανακατασκευασμένων 3D αντικειμένων



Scan 65 Κρανίο	Chamfer Distance(mm)			
	STL DTU scans	Camp MVS algorithm	Furu MVS algorithm	Tola MVS algorithm
Μέθοδος Εξαγωγής GT Mesh				
Positional Encoding	0.971	1.27	1.70	2.36
MR HashGrid3D	1.13	1.22	1.72	2.13
NFFB	1.00	1.29	1.75	2.24
StylemodNFFB	0.957	1.2	1.65	2.14
StylemodNFFB (TinyCudaNN)	1.11	1.48	1.99	2.44

Αποτελέσματα αξιολόγησης μετρικής Chamfer Distance(DTU Scene:
Skull[65])

Chamfer Distance

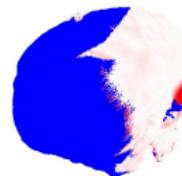
Απεικόνιση Chamfer Distance - Σκηνή 65 - StylemodNFFB



34



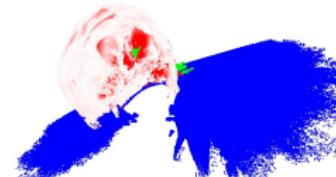
d2s front



d2s back



s2d front



s2d back

Συμβολή

Συστηματική Έρευνα Κωδικοποίησης

Επιτάχυνση και Βελτίωση Απόδοσης 3D ανακατασκευής



35



Positional NTK

MR HashGrid 3D

NFFB

Stylemod NFFB

Stylemod NFFB(tcnn)

Εκπαίδευση 100 εποχών

Προβλήματα

Αστοχίες Ανακατασκευών, Χρόνος Εκπαίδευσης, Απαιτούμενοι Πόροι



36

- Η πλήρης εκπαίδευση (2000 εποχές) για κάθε σκηνή διαρκεί κατά μέσο όρο 20 ώρες (παρά την επιτάχυνση στα μοντέλα TCNN - 14ώρες)
- Οι υπολογιστικοί πόροι σε μνήμη κατά την εκπαίδευση είναι μεγάλοι και κατά την ανάκληση στην αξιολόγηση απαιτείται αρκετό VRAM (Marching Cubes Resolution)
- Συντηρητικές τιμές σε HashGrid resolution levels λόγω απαιτήσεων σε πόρους
- Artifacts και Αστοχίες Ανακατασκευών όταν υπάρχει σύνθετο πεδίο φωτισμού σε ορισμένα μοντέλα.



Solved with StylemodNFFB





Μελλοντικές Εφαρμογές

- Εφαρμογή μοντέλων κωδικοποίησης σε άλλα δίκτυα που εφαρμόζουν και ογκομετρική αποτύπωση (NeuS, NeRF) ή σε δίκτυα ανακατασκευής εικόνας, ή βίντεο.
- Αφαίρεση επίβλεψης μάσκας και χρήση δεδομένων από μονοσκοπικό βίντεο για άμεση ανακατασκευή χώρου
- Χρήση σε δίκτυα με CNF (Conditional Neural Fields) τα οποία χρησιμοποιούνται από GAN δίκτυα για παραγωγή 3D περιεχομένου (π.χ. παραμορφώσεις για μηχανές προσομοίωσης)

Προϋποθέσεις Low-Level Code , Refine CUDA kernels

- Μεταφορά κώδικα εξ' ολοκλήρου σε χαμηλότερου επιπέδου γλώσσα (C++) για εκμετάλλευση καλύτερα της παραλληλοποίησης και αποφυγή memory leaks
- Βελτιστοποίηση των ήδη υπάρχων μεθόδων σε CUDA και χρήση mixed precision training

Ενδεικτική Βιβλιογραφία



38

Διαδικασία Ανακατασκευής

Μεταβίβαση παραμέτρων εμφάνισης
(γ του Rendering Network)

Σας Ευχαριστώ, Ερωτήσεις;