



HUST

TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

ONE LOVE. ONE FUTURE.



TRƯỜNG ĐẠI HỌC
BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Sử dụng đồng thời Ước tính độ sâu và Kỹ thuật hình ảnh Stereo để mô phỏng hình ảnh rẽ Cây

Nguyễn Cảnh Anh 20202291

ONE LOVE. ONE FUTURE.

NỘI DUNG



- Giới thiệu



- Kiến thức liên quan



- Nội dung phương pháp



- Thực nghiệm



- Kết luận

ONE LOVE. ONE FUTURE.

I. Giới thiệu

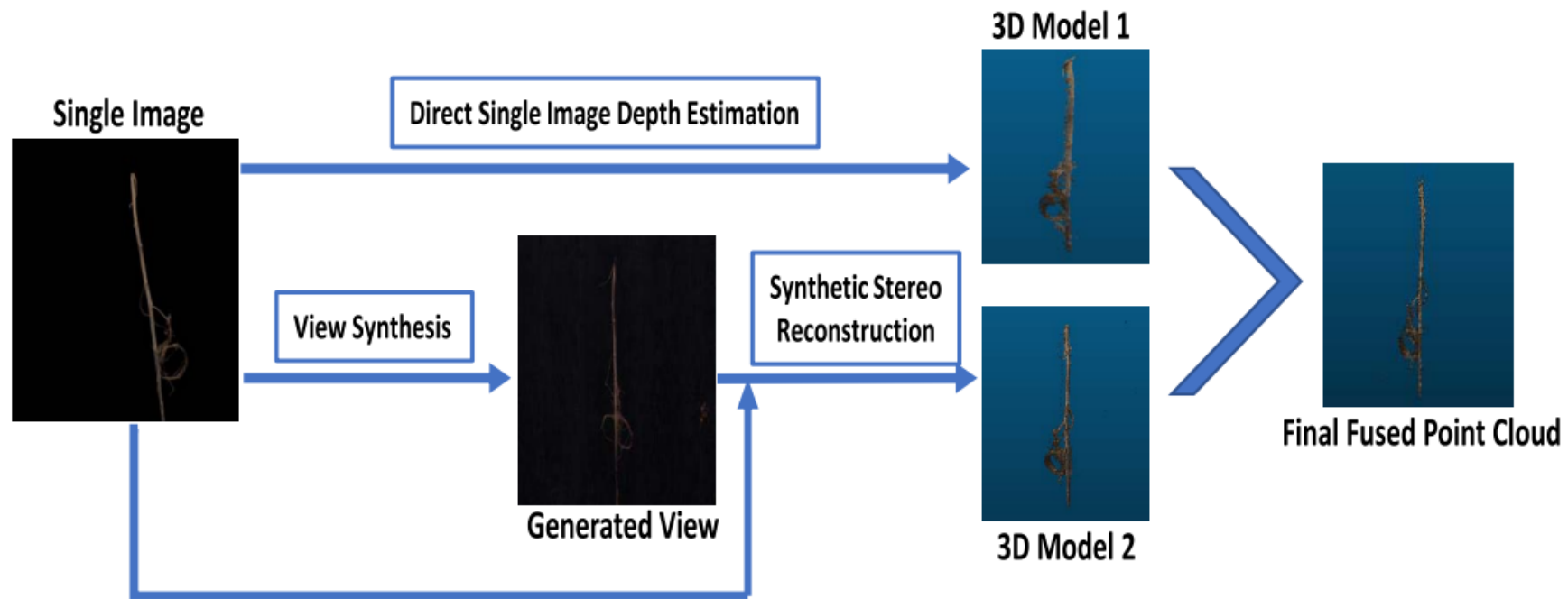
Rễ cây có vai trò quan trọng trong sự phát triển và sinh trưởng của cây.

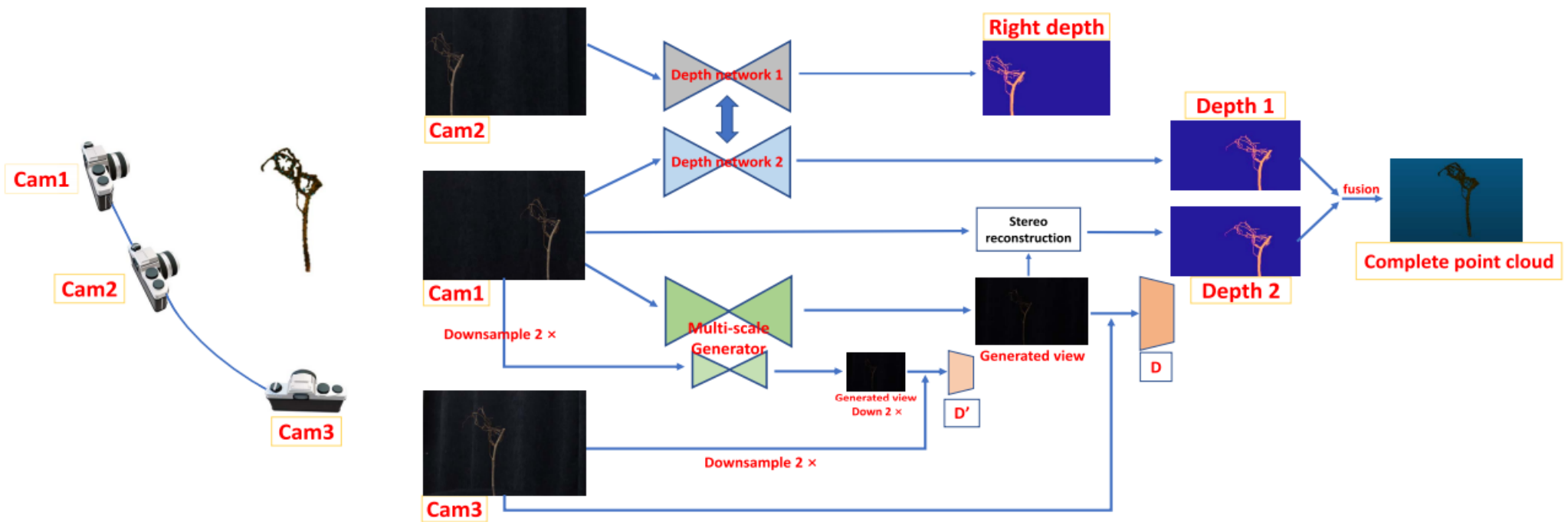
Các mô hình hiện nay như [structure-from-motion](#) và [shape-from-silhouette](#) cần nhiều hình ảnh đầu vào và phức tạp.

Họ đưa ra một mô hình học máy không giám sát (unsupervised learning) để ước tính độ sâu hình ảnh rễ cây từ một ảnh duy nhất, từ đó xây dựng lại hình ảnh 3D.

Để khắc phục những lỗi trong quá trình thực hiện, họ đồng thời sử dụng mạng GAN và tái tạo ảnh Stereo để mô phỏng theo một góc nhìn khác. Sau đó cho ra kết quả cuối cùng.

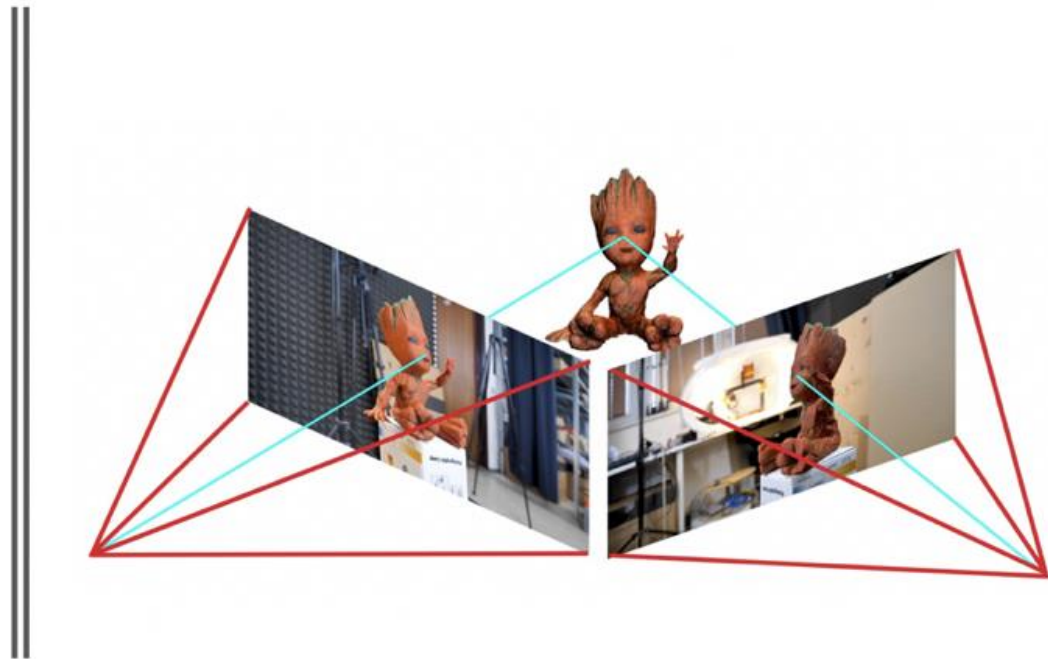
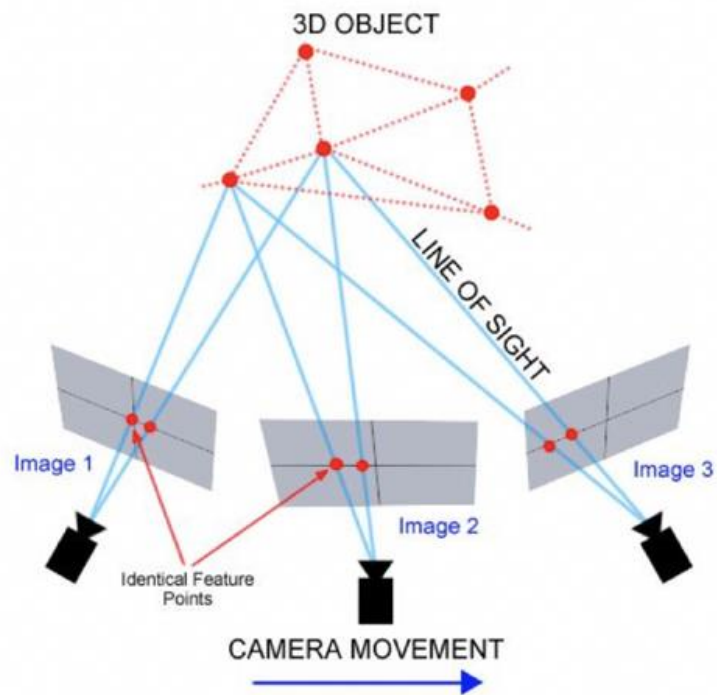
Mô hình





II. Kiến thức liên quan

A. Tái tạo đa góc



II. Kiến thức liên quan

B. Mô hình tái cấu trúc có giám sát

- Phương pháp sử dụng học máy để xây dựng lại một đối tượng 3D.
- Mô hình học máy được huấn luyện trên một tập dữ liệu huấn luyện có sẵn
- Khi đã được huấn luyện, mô hình có thể sử dụng để dự đoán cấu trúc 3D của đối tượng hoặc cảnh vật từ các hình ảnh mới.
- Phương pháp này được sử dụng trong nhiều ứng dụng, bao gồm thị giác máy tính, thực tế ảo, truyền thông và robot học.

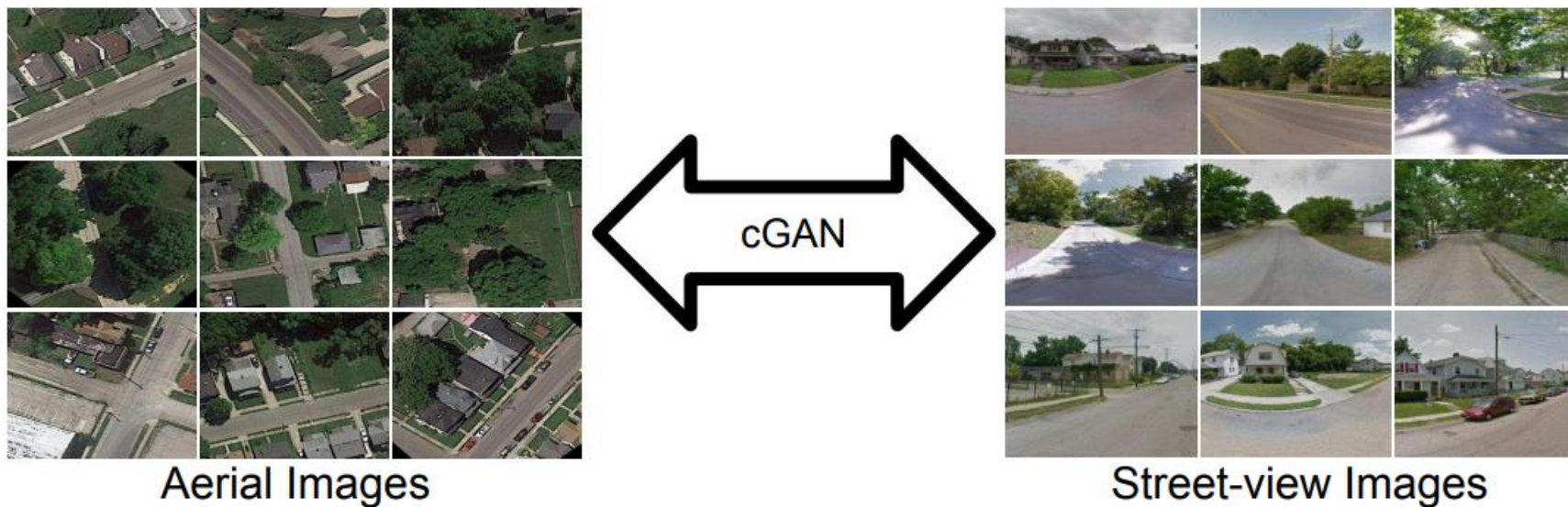
II. Kiến thức liên quan

C. Mô hình tái cấu trúc không giám sát

- Phương pháp sử dụng học máy không giám sát để xây dựng lại một đối tượng 3D.
- Mô hình học máy được huấn luyện để tự động tìm ra các đặc trưng quan trọng trong dữ liệu hình ảnh và ánh xạ chúng sang không gian 3D.
- Khi đã được huấn luyện, mô hình có thể sử dụng để dự đoán cấu trúc 3D của đối tượng hoặc cảnh vật từ các hình ảnh mới mà không cần thông tin về cấu trúc 3D trước.
- Tuy nhiên, phương pháp này thường đòi hỏi thời gian huấn luyện và sự khó khăn trong việc đánh giá kết quả.

II. Kiến thức liên quan

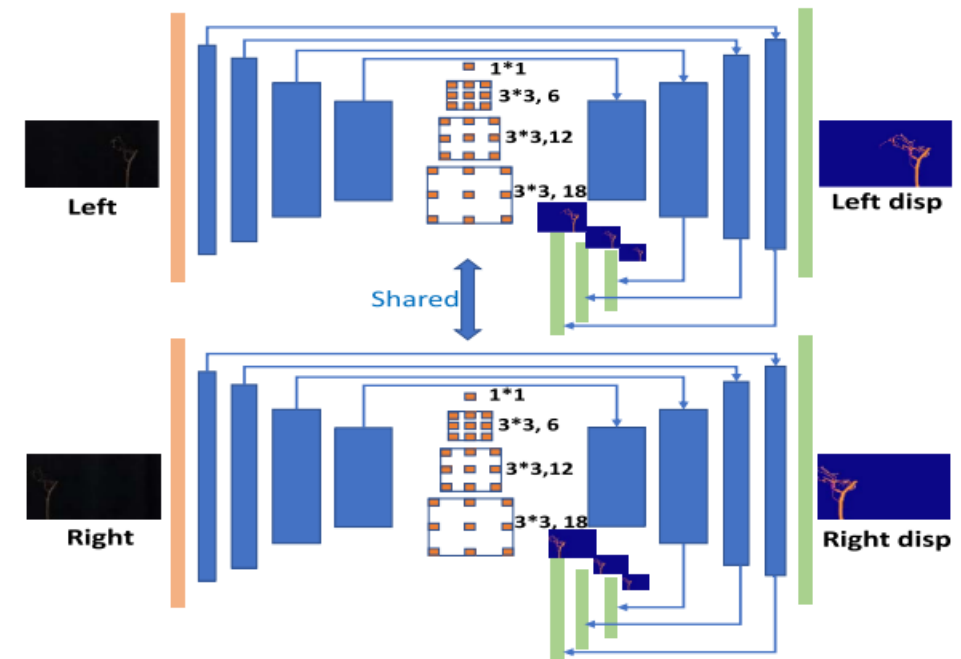
D. Tổng hợp hình ảnh nhiều góc nhìn



III. Nội dung phương pháp

A. Ước tính độ sâu trực tiếp

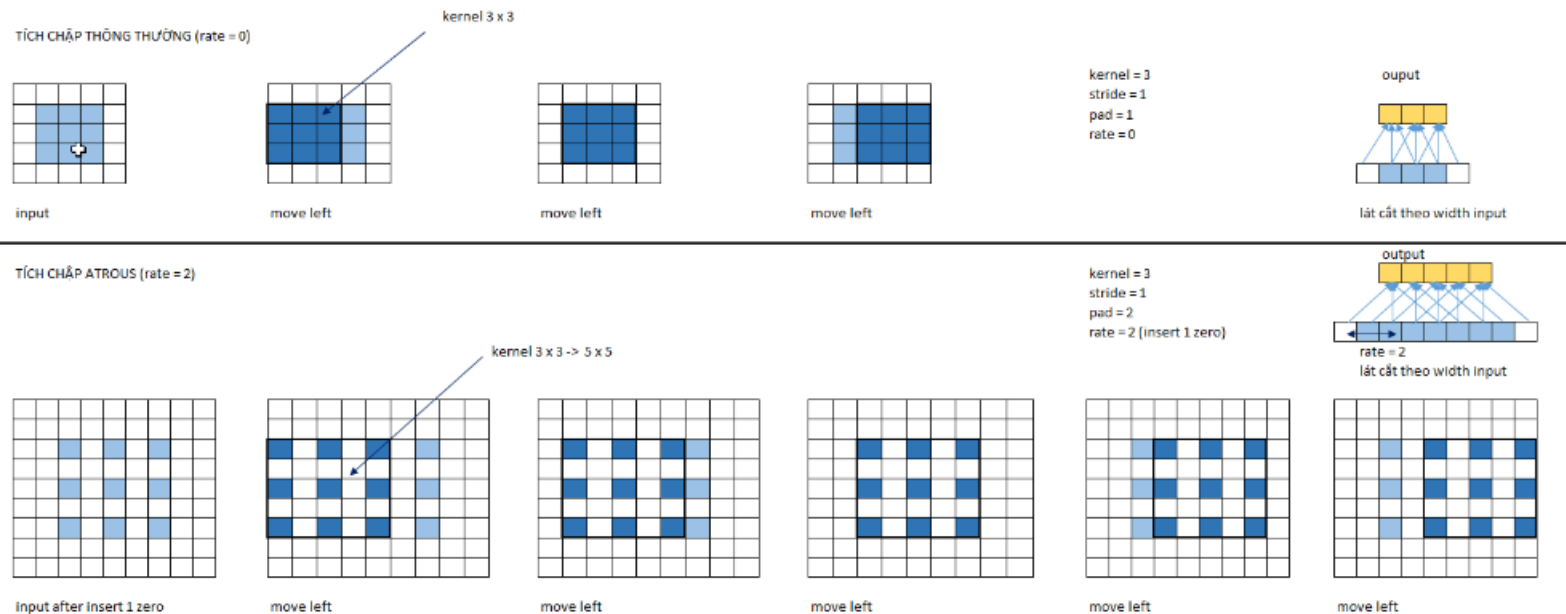
- Mô hình học máy không giám sát
- Sử dụng 2 mạng học tích chập và chia sẻ trọng số cho nhau.
- Sử dụng Tích chập ASPP trong 2 mạng tích chập này.



III. Nội dung phương pháp

A. Ước tính độ sâu trực tiếp

Tích chập ASPP



III. Nội dung phương pháp

A. Ước tính độ sâu trực tiếp

Hàm tổn thất thông tin:

$$\begin{aligned} L_{\text{consis}} = & \frac{1}{N} \sum_{ij} \frac{a}{2} (1 - SSIM(I_{ij}, \text{Warp}(I_{ij}))) \\ & + \frac{2}{3} (1 - a) (\|I_{ij} - \text{Warp}(I_{ij})\|_1) \\ & + \frac{1}{3} (1 - a) (\|I_{ij} - \text{Warp}(I_{ij})\|_2) \end{aligned}$$

I_{ij} : Đại diện cho 2 đầu vào

$\text{Warp}(I_{ij})$: đại diện cho hình ảnh bị biến dạng với bản đồ chênh lệch bằng cách di chuyển các pixel từ hình ảnh đầu vào ban đầu dọc theo đường biểu cực

$a = 0.85$

III. Nội dung phương pháp

A. Ước tính độ sâu trực tiếp

$$L_{\text{keypoints}} = \sum_i \sum_{j \in N_i} \left\| (K P_i - K P_j) - (K \hat{P}_i - K \hat{P}_j) \right\|^2$$

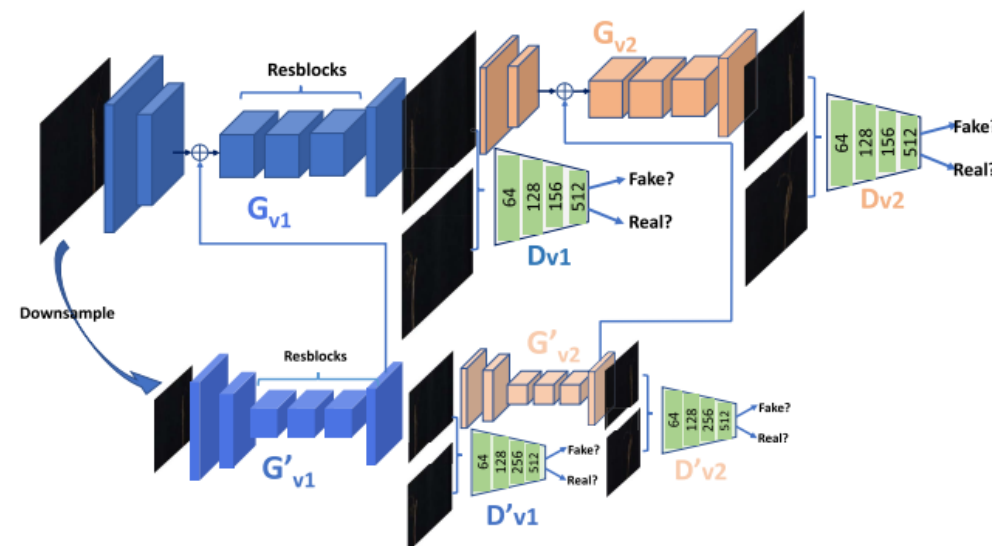
$$L_{\text{edge}} = \sum_{ij} \text{Hamm_Dist} (\|E_{ij} - \text{Warp}(E_{ij})\|)$$

$$L_{\text{key_edge}} = \sum \alpha \times L_{\text{edge}} + (1 - \alpha) L_{\text{keypoints}}$$

III. Nội dung phương pháp

B. Kỹ thuật ảnh Stereo dựa vào tổng hợp xem chéo

- Tổng hợp các hình ảnh ở góc nhìn khác sử dụng mạng GAN
- GAN gồm 2 thành phần chính là bộ tạo và bộ phân biệt.
- Bộ tạo đóng vai trò tạo ra một hình ảnh khác từ hình ảnh ban đầu ở góc nhìn khác
- Bộ phân biệt có nhiệm vụ đánh giá xem ảnh đó có đạt chất lượng hay không



III. Nội dung phương pháp

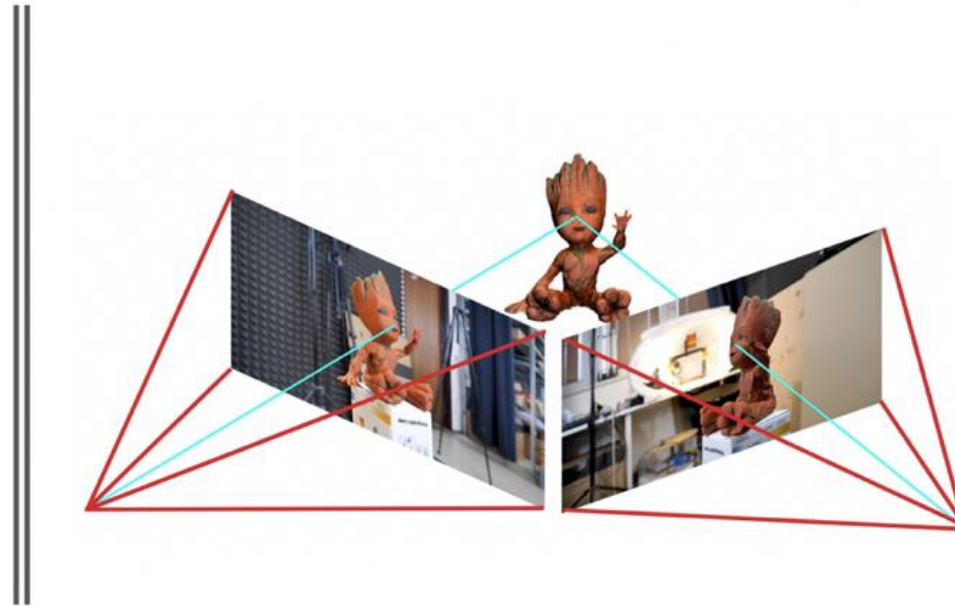
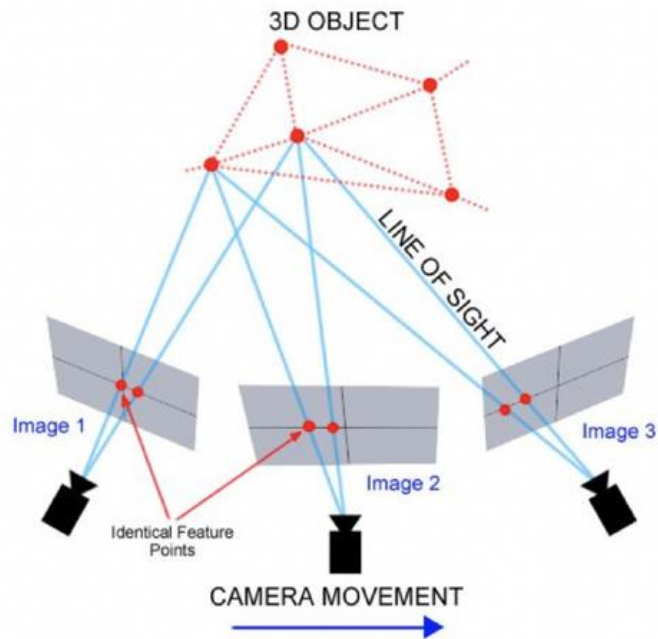
B. Kỹ thuật ảnh Stereo dựa vào tổng hợp xem chéo

$$L_{Adv, v1 \rightarrow v2} = \sum_{l=1}^L E_{t \in p_{v1}(c)} \left[1 - \log \left(D(c)^{(l)} \right) \right] + E_{t \in p_{v2}(t)} \left[\log D \left(G(t)^{(l)} \right) \right]$$

$$L_{cycle} = \sum_{l=1}^L E_{c \in Ori(c)} \left[\left\| G_{v2}^{(l)} \left(G_{v1}^{(l)}(c) \right) - c \right\|_1 \right]$$

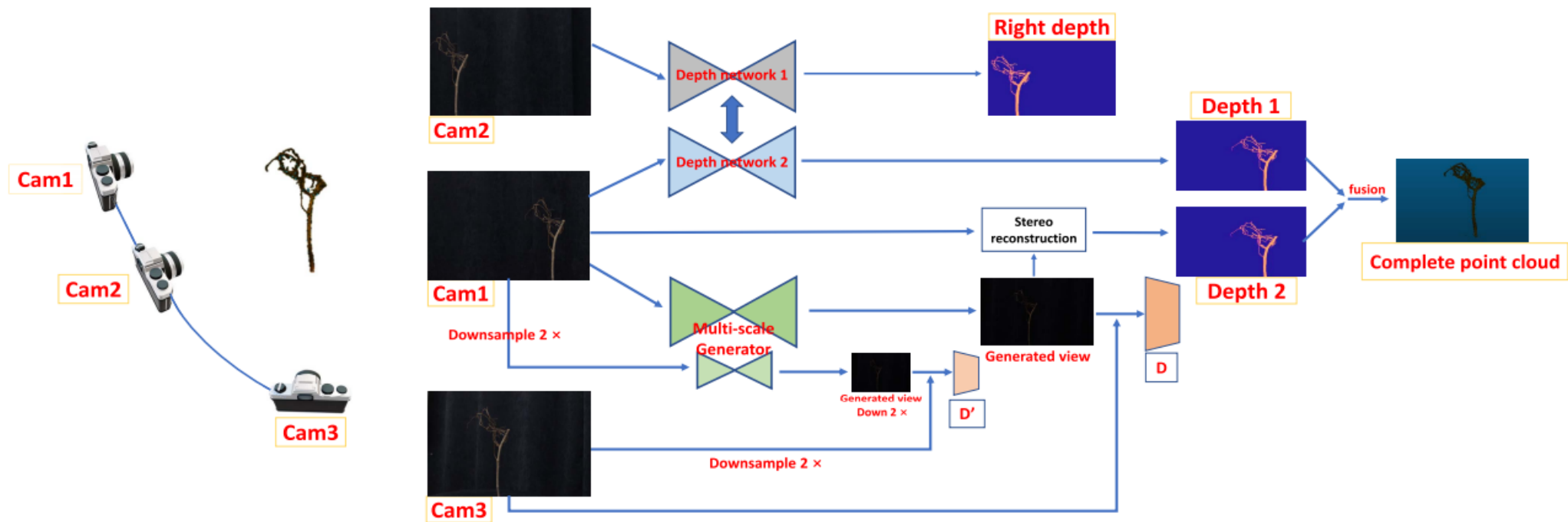
III. Nội dung phương pháp

B. Kỹ thuật ảnh Stereo dựa vào tổng hợp xem chéo



III. Nội dung phương pháp

B. Kỹ thuật ảnh Stereo dựa vào tổng hợp xem chéo



IV: Thực nghiệm

1. Thu thập tập dữ liệu

Một bộ dữ liệu thế giới thực có độ phân giải cao được chụp bởi ba máy ảnh Canon EOS DSLR, có tên là “Bộ dữ liệu gốc DSLR”

Ba máy ảnh (CANON ÉO REBEL SL1)

Sensor Name	Canon EOS SL1
Sensor Type	CCD
Pixel Dimension	2592 × 1728
Length of Focus	20 mm
Sensor Size	22.3 × 14.9mm
FOV	58 × 41

IV: Thực nghiệm



Tập dữ liệu gốc được thu thập với mỗi hàng ảnh được chụp từ một trong 3 điểm quan sát khác nhau.

2. Mô hình Training

- Dự đoán chế độ xem chéo, trước tiên họ chia tỷ lệ hình ảnh thành 720×480 .
- Sau khi training, chế độ xem nguồn và chế độ xem dự đoán sẽ được áp dụng kỹ thuật hình ảnh Stereo nhằm dự đoán bản đồ độ sâu nhiều chế độ xem.
- Sau đó, họ sử dụng các bản đồ độ sâu để tạo ra các đám mây điểm 3D gốc.

3. Mô phỏng ảnh rẽ cây 3D

Họ sử dụng phân tích thử nghiệm để đánh giá các mô hình của mình dựa trên các chỉ số sau: PSNR và SSIM để đánh giá hiệu suất tổng hợp chế độ xem; Điểm thể lực và Khoảng cách trung bình để đánh giá hiệu suất của mô hình 3D được xây dựng lại.

$$PSNR = 10\log_{10}\left(\frac{MAX^2}{MSE}\right).$$

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1) + (2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

4. So sánh với các phương pháp hiện đại trước.

A) So sánh định tính: Để có kết quả định tính phương pháp của họ được đánh giá trực quan để dự đoán chế độ xem bằng các phương pháp gần đây nhất: Pix2Pix và SingleGAN.

B) So sánh định lượng: So sánh và phân tích số bao gồm PSNR và SSIM trong mạng dự đoán chế độ xem bằng phương pháp của họ và Khoảng cách trung bình với giá trị phương sai giữa các đám mây điểm cho hiệu suất tái tạo 3D.

IV: Thực nghiệm

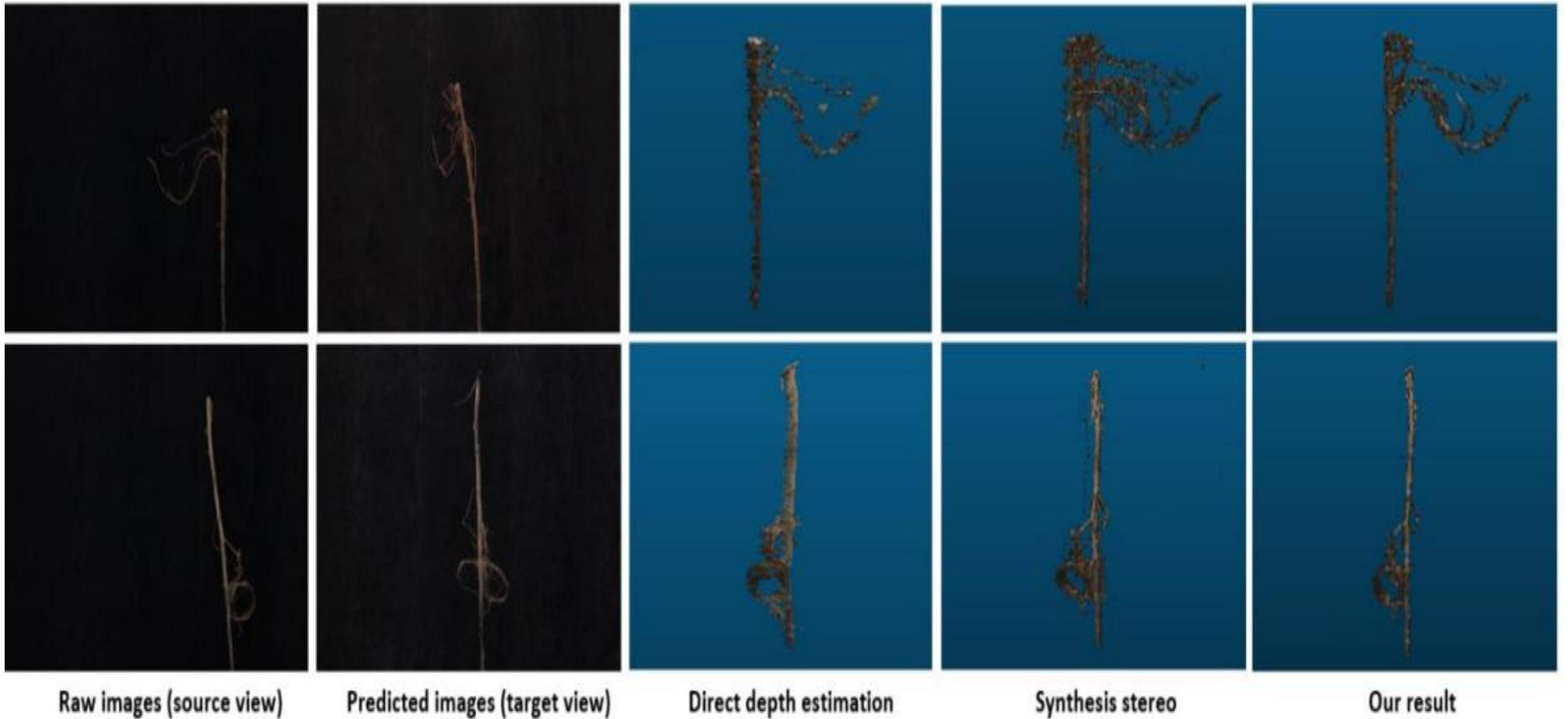
Điểm PSNR và SSIM trung bình của chế độ xem mới được đề xuất phương pháp tổng hợp so với dựa trên Gan gần đây

Method	PSNR	SSIM
Pix2Pix [45]	20.4	0.80
SingleGAN [39]	18.8	0.83
DualGAN [46]	23.6	0.85
Cross-net [30]	21.9	0.82
Pix2PixHD [47]	24.7	0.86
Ours (w/o multi-scale structure)	23.8	0.85
Ours (full pipeline)	25.9	0.87

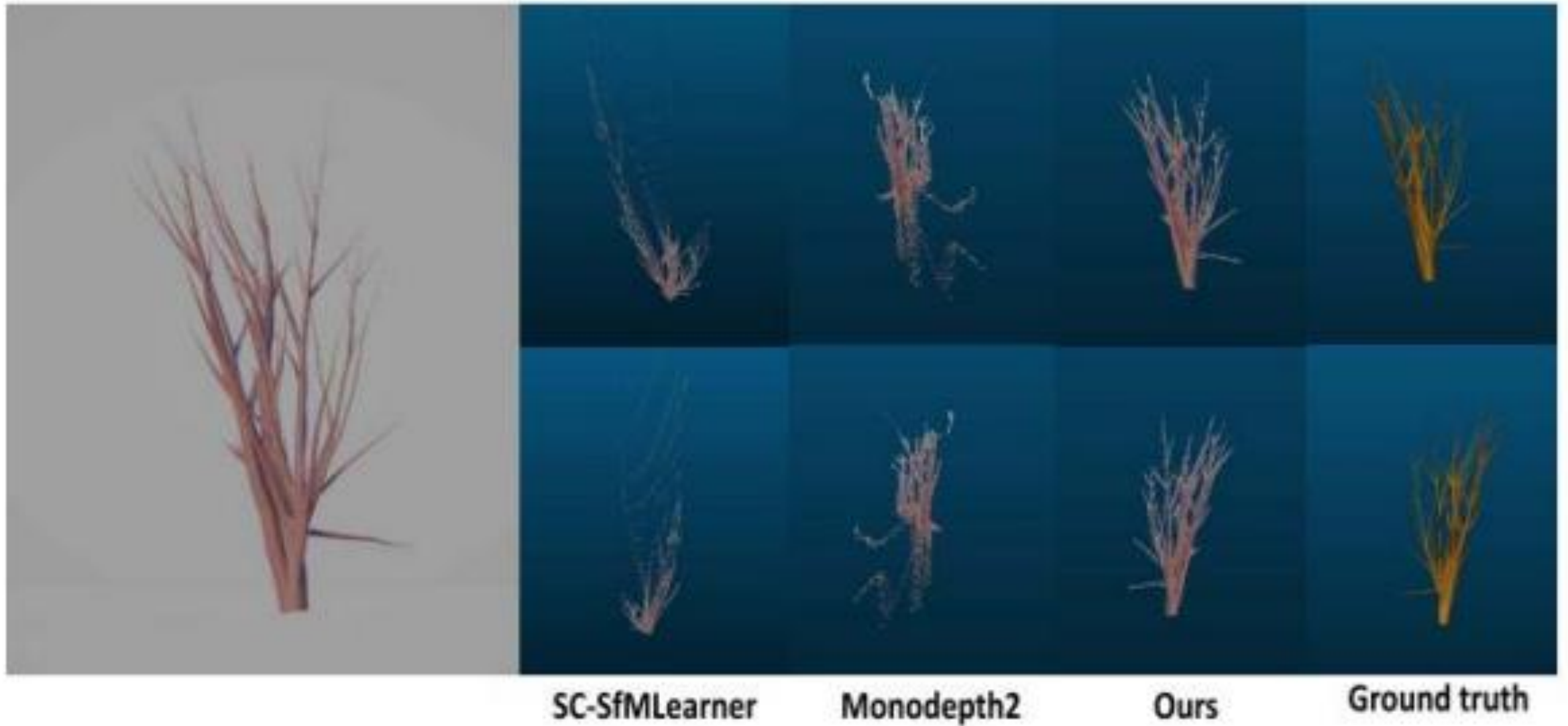
Khoảng cách trung bình và phương sai phương pháp của họ so với các phương pháp khác

Method	Mean Distance	Variance
Monodepth [26]	109.52	38.61
Monodepth2 [33]	37.69	12.47
Un-stereo-depthGAN [43]	66.78	15.76
MVDepthNet [17]	58.43	13.14
SC-SfMLearner [44]	59.63	14.10
Our direct depth estimation	35.92	12.65
Our synthetic stereo	33.26	9.79
Our full pipeline	29.42	9.23

IV: Thực nghiệm



IV: Thực nghiệm

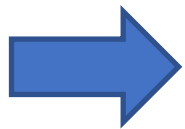
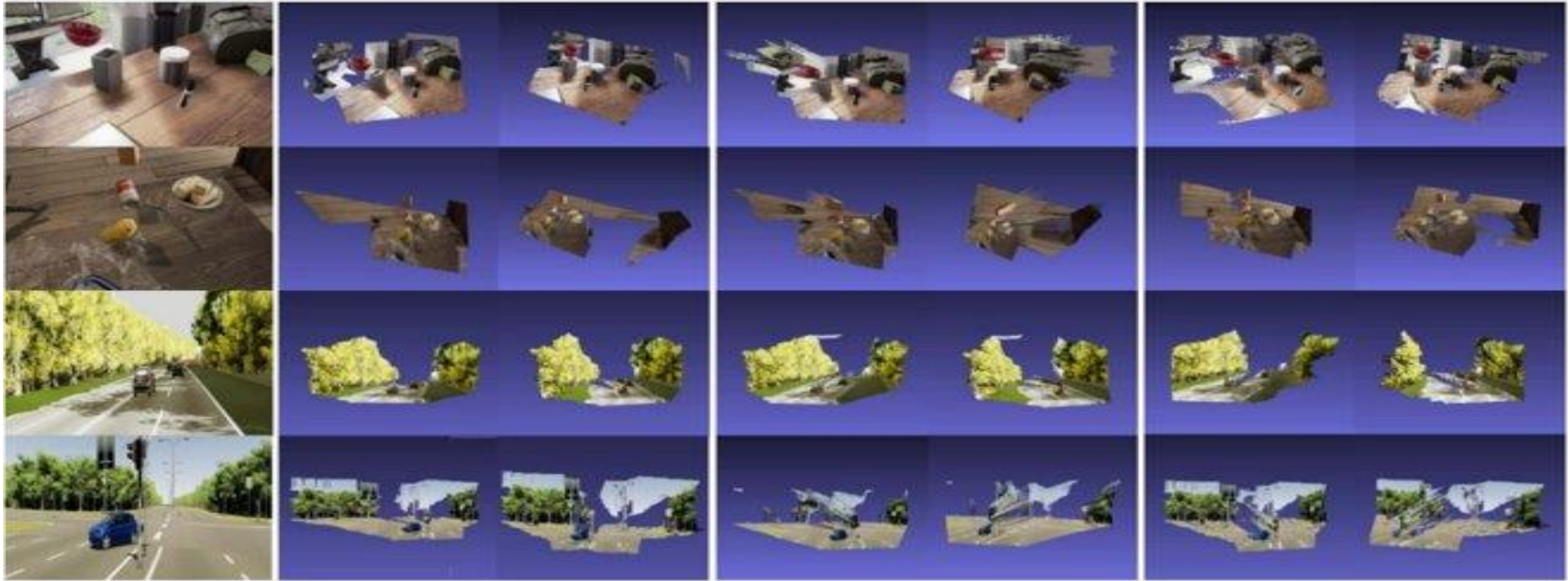


IV: Thực nghiệm

Khoảng cách trung bình và phương sai phương pháp của họ so với các phương pháp khác

Method	Mean Distance	Variance
Monodepth2 [33]	33.47	14.16
SC-SfMLearner [44]	47.12	13.01
Our direct depth estimation	30.98	12.23
Our synthetic stereo	28.62	10.81
Our full pipeline	24.30	10.24

IV: Thực nghiệm



Phương pháp của họ có thể tạo ra ít biến dạng và kéo dài nhất trong các hình dạng và khung của họ có thể bảo toàn nhiều chi tiết hơn cho các nhánh mỏng.



- Trong bài báo này, họ đã trình bày một phương pháp mới để mô phỏng hình ảnh rễ cây 3D chính xác từ một hình ảnh đầu vào duy nhất.
- Phương pháp được đề xuất vượt trội đáng kể so với các phương pháp tiên tiến nhất trong việc tái cấu trúc các rễ cây mỏng và phức tạp từ một hình ảnh duy nhất.
- Mặc dù được thiết kế đặc biệt cho rễ cây, nhưng phương pháp được đề xuất vẫn có thể áp dụng chung cho các cảnh phổ biến khác như đồ vật trong nhà và cảnh ngoài trời.

A large graphic on the left side of the slide. It features a dark blue background with a circular pattern of red dots of varying sizes, creating a sense of depth and movement. The word "HUST" is centered within this graphic in a bold, white, sans-serif font.

HUST

THANK YOU !