

# Single View Stereo Matching

Hồ Thịnh - 19522274

# Thông tin

- Lớp: CS519.M11.CLC
- Link Github: <https://github.com/bou24-1/CS519.M11.KHCL>
- Link YouTube video:
- Người thực hiện: Hồ Thịnh - 19522274

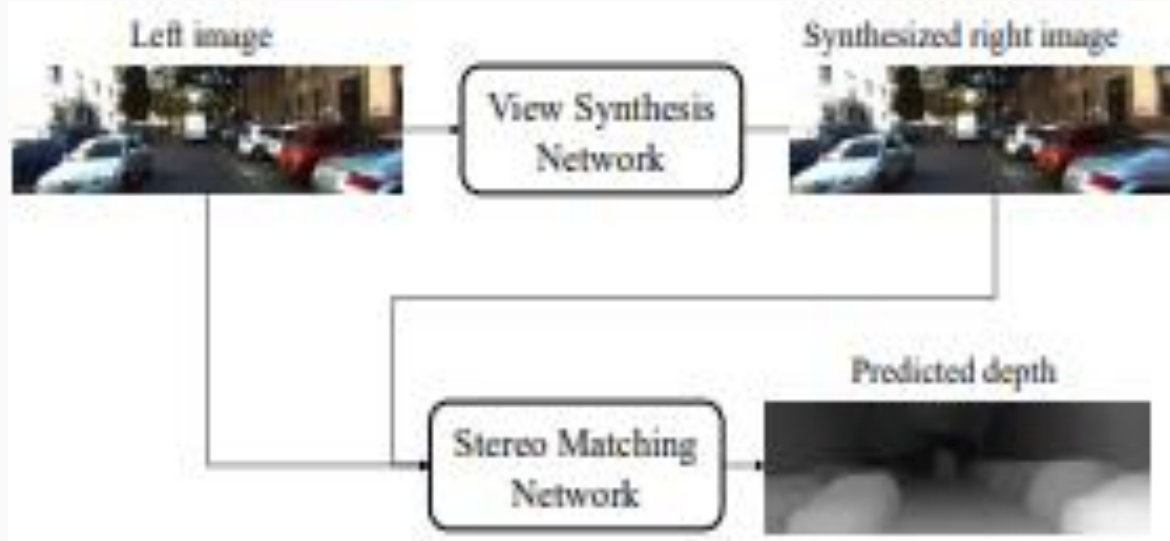


# Tóm tắt

1. Giới thiệu
2. Mục tiêu
3. Nội dung và phương pháp
4. Kết quả dự kiến

# Giới thiệu

- Bài toán Monocular Depth Estimation
- Hạn chế trong các cách tiếp cận trước đây:
  - Sự khó khăn trong việc học thông tin ngữ nghĩa trong các mô hình.
  - Phụ thuộc vào dữ liệu được gán nhãn để thiết lập mối quan hệ giữa cảnh và độ sâu.



# Mục tiêu

- Trong bài báo này chúng tôi muốn đề xuất một cách tiếp cận để giải quyết bài toán Monocular Depth Estimation và cho thấy bài toán này có thể được xây dựng và giải quyết như một Stereo Matching Problem.
- Chúng tôi muốn cho thấy:
  - Cả hai mạng neuron chúng tôi xây dựng đều tuân theo/ phù hợp với nguyên tắc hình học.
  - Mô hình có thể được đào tạo mà không yêu cầu về dữ liệu về độ sâu thực (label) nhưng vẫn hoạt động tốt.
  - Toàn bộ hệ thống được đào tạo end-to-end giúp tối ưu hoá sự chính xác về mặt hình học.
- Về độ chính xác, cho thấy sự vượt trội của phương pháp này so với Stereo Blocking Matching Algorithm.

# Nội dung và Phương pháp

- Chia bài toán Monocular Depth Estimation thành 2 bài toán con là View Synthesis Problem và Stereo Matching Problem.
- Cài đặt các mạng học sâu để giải quyết cho từng bài toán con. Sau đó, toàn bộ pipeline được huấn luyện end-to-end.

# Nội dung và Phương pháp

## Nội dung 1: Cài đặt View Synthesis Network

- Áp dụng thiết kế của mô hình Deep3D và phát triển mạng học sâu của chúng tôi dựa trên nó.
- Tại sao lại là Mô hình Deep3D?
  - Mô hình đề xuất một lược đồ xác suất mới để chuyển các pixel từ hình ảnh ban đầu thay vì ước tính độ chính xác về mặt hình học như các phương pháp warp-based trước đây.
  - Trực tiếp tạo ra sự biến đổi từ hình ảnh trái sang hình ảnh bên phải bằng cách sử dụng một lớp lựa chọn có thể phân biệt. Điều này phù hợp với binocular views được đồng bộ hoá, hiệu chỉnh từ camera.

# Nội dung và Phương pháp

## Nội dung 2: Xây dựng Stereo Matching Network

- Nhằm đơn giản hoá trong tính toán, Stereo Matching Network của chúng tôi chuyển đổi vấn đề hiểu cảnh cấp cao thành vấn đề đối sánh 1D
- Để sử dụng tốt hơn mối quan hệ hình học giữa hai chế độ xem, chúng tôi lấy ý tưởng về mối tương quan 1D được sử dụng trong DispNetC. Chúng tôi tiếp tục áp dụng cấu trúc DispFulNet để đạt được dự đoán có độ phân giải đầy đủ.



# Kết quả dự kiến

Thông qua kết quả về độ chính xác cũng như thực nghiệm:

- Cho thấy cách tiếp cận của chúng tôi loại bỏ các hạn chế từ các cách tiếp cận trước.
- Cho thấy cả hai mạng đều có thể bảo toàn chức năng ban đầu của chúng sau khi đào tạo end-to-end.
- Chúng tôi sẽ so sánh với kết quả của các cách tiếp cận trước đây để cho thấy cách tiếp cận này thực sự hiệu quả.

# Tài liệu tham khảo

- [1] W. Chen, Z. Fu, D. Yang, and J. Deng. Single-image depth perception in the wild. In NIPS, 2016.
- [2] Y. Kuznietsov, J. Stuckler, and B. Leibe. Semi-supervised deep learning for monocular depth map prediction. 2017.
- [3] C. H. Esteban, G. Vogiatzis, and R. Cipolla. Multiview photometric stereo. TPAMI, 30(3):548–554, 2008.
- [4] R. Ranftl, V. Vineet, Q. Chen, and V. Koltun. Dense monocular depth estimation in complex dynamic scenes. In CVPR, 2016.
- [5] J. Xie, R. Girshick, and A. Farhadi. Deep3d: Fully automatic 2d-to-3d video conversion with deep convolutional neural networks. In ECCV, 2016.