**Các thành phần chính**: Bài báo nói về Stripdet, một framework nhẹ đột phá được thiết kế để đạt hiệu quả trên thiết bị có tài nguyên không quá lớn.Các thử nghiệm rộng rãi xác nhận sự vượt trội của StripDet. Với chỉ 0,65M tham số, mô hình của StripDet đạt 79,97% cho việc phát hiện ô tô, vượt trội hơn mô hình khác. Hơn nữa, StripDet vượt trội hơn các phương pháp nhẹ và dựa trên chưng cất tri thức gần đây, đạt được sự cân bằng tối ưu giữa độ chính xác và hiệu quả đồng thời khẳng định mình là một giải pháp thực tế cho việc phát hiện 3D trong thế giới thực trên các thiết bị biên.

1. Đề xuất Strip Attention Block (SAB) – đổi mới cốt lõi

SAB sử dụng tích chập dải bất đối xứng (strip convolution) thay cho tích chập 2D chuẩn.Thiết kế này cho phép nắm bắt phụ thuộc không gian dài hạn (long-range spatial dependencies) trong dữ liệu đám mây điểm.  
Chi phí tính toán được giảm mạnh, từ bậc hai O(K²) xuống tuyến tính O(K).  
Nhờ đó, mô hình vẫn giữ khả năng biểu diễn mạnh mẽ trong khi trở nên nhẹ và hiệu quả.  
 2. Xây dựng StripDet – kiến trúc phát hiện 3D nhẹ, end-to-end

Tái cấu trúc backbone theo hướng thân thiện phần cứng bằng cách tích hợp SAB với:

* 1. Depthwise separable convolution để giảm FLOPs.
  2. Chiến lược hợp nhất đa tỉ lệ đơn giản (upsampling + concatenation) thay cho các module phức tạp như FPN.

Cấu trúc tổng thể gồm:

* 1. Pillar Encoder → chuyển point cloud sang bản đồ BEV.
  2. Backbone phân cấp (dựa trên SAB) → trích xuất đặc trưng.
  3. Detection Head → ba nhánh cho phân loại, hồi quy hộp và dự đoán hướng.

Cách thiết kế này giúp StripDet vừa hiệu quả đầu-cuối, vừa phù hợp để triển khai trên thiết bị biên.

3. Hiệu năng vượt trội trên thực nghiệm

Trên tập KITTI, StripDet chỉ có 0.65 triệu tham số nhưng đạt 79.97% mAP cho phát hiện xe hơi. So với baseline PointPillars: Giảm tham số gấp 7 lần và FLOPs gấp 3.6 lần. Nhưng vẫn tăng độ chính xác. So với các phương pháp lightweight khác (LW-S-3) và các phương pháp dựa trên knowledge distillation (PointDistiller). StripDet nhẹ hơn, huấn luyện đơn giản hơn. Nhưng kết quả lại ổn định và vượt trội hơn trên cả ba lớp (xe, người đi bộ, xe đạp).  
Kết luận: StripDet đạt được cân bằng tối ưu giữa accuracy và efficiency, thích hợp cho ứng dụng thực tế thời gian thực.

**Bài văn:**

### **StripDet: Kiến Trúc Gọn Nhẹ và Hiệu Quả Cho Nhận Diện Vật Thể 3D**

Trong lĩnh vực nhận diện vật thể 3D từ dữ liệu đám mây điểm (point cloud), một thách thức lớn là các mô hình hiện đại dù đạt độ chính xác cao nhưng lại đòi hỏi chi phí tính toán (computational cost) và dung lượng bộ nhớ (memory footprint) khổng lồ. Điều này tạo ra rào cản lớn khi triển khai chúng trên các thiết bị biên (edge devices) có tài nguyên hạn chế. Các phương pháp tối ưu hóa truyền thống như cắt tỉa (pruning), lượng tử hóa (quantization) hay chưng cất tri thức (knowledge distillation) tuy đã được áp dụng nhưng đều bộc lộ những nhược điểm cố hữu. Pruning và quantization có nguy cơ làm suy giảm khả năng biểu diễn của mô hình, trong khi knowledge distillation yêu cầu một quy trình huấn luyện phức tạp giữa mô hình "thầy" và "trò", gây tốn kém và kém linh hoạt.

Để giải quyết triệt để những vấn đề này, bài báo giới thiệu **StripDet**, một framework được thiết kế gọn nhẹ ngay từ cấp độ kiến trúc. Đóng góp cốt lõi và đột phá nhất của nghiên cứu là **Khối Chú Ý Dạng Dải (Strip Attention Block - SAB)**. Thay vì sử dụng các phép tích chập 2D tiêu chuẩn vốn tốn kém, SAB sử dụng các phép tích chập dạng dải bất đối xứng (1×K và K×1). Thiết kế thông minh này cho phép mô hình nắm bắt hiệu quả các mối tương quan không gian tầm xa (long-range spatial dependencies) trong dữ liệu với chi phí tính toán chỉ tăng tuyến tính O(K), một cải tiến vượt bậc so với chi phí bậc hai O(K2) của các phương pháp truyền thống. SAB được cấu trúc như một khối thặng dư (residual block), kết hợp khả năng thu nhận thông tin toàn cục và duy trì năng lực học các đặc trưng cục bộ một cách hiệu quả.

Kiến trúc tổng thể của StripDet bao gồm ba thành phần chính. Đầu tiên, **Bộ mã hóa Trụ (Pillar Encoder)** biến đổi dữ liệu point cloud thưa thớt thành một bản đồ đặc trưng có cấu trúc theo góc nhìn từ trên xuống (Bird’s-Eye View). Tiếp theo, **Mạng Xương Sống Phân Cấp (Hierarchical Backbone)**, trái tim của mô hình, tận dụng sức mạnh của các khối SAB, kết hợp với tích chập tách biệt theo chiều sâu (depthwise separable convolution) và cơ chế hợp nhất đa tỷ lệ đơn giản để tối ưu hóa hiệu suất mà vẫn giữ được sự tinh gọn. Cuối cùng, **Đầu Mạng Nhận Diện (Detection Head)** với ba nhánh song song thực hiện các tác vụ phân loại, hồi quy hộp giới hạn và dự đoán hướng của vật thể. Toàn bộ mô hình được huấn luyện end-to-end thông qua một hàm mất mát đa tác vụ, sử dụng trình tối ưu hóa AdamW và chiến lược tốc độ học one-cycle.

Kết quả thực nghiệm trên bộ dữ liệu KITTI đã chứng minh hiệu năng ấn tượng của StripDet. Với chỉ **0.65 triệu tham số**, mô hình vẫn đạt được **79.97% mAP** (mean Average Precision) trong tác vụ nhận diện xe hơi. So với mô hình nền PointPillars, StripDet giảm số lượng tham số tới 7 lần và FLOPs 3.6 lần nhưng vẫn giữ được độ chính xác cạnh tranh. Khi đặt cạnh các mô hình gọn nhẹ khác hay các phương pháp chưng cất tri thức như PointDistiller, StripDet không chỉ đơn giản hơn trong huấn luyện mà còn cho thấy sự cân bằng vượt trội giữa độ chính xác và hiệu suất tính toán.

Tóm lại, StripDet đã chứng tỏ mình là một kiến trúc hiệu quả nội tại, mang đến giải pháp tối ưu giữa độ chính xác và tài nguyên mà không cần đến các quy trình huấn luyện phức tạp. Thành công này mở ra tiềm năng to lớn cho các ứng dụng nhận thức 3D thời gian thực trên thiết bị biên, đặc biệt trong các lĩnh vực tiên phong như xe tự hành, robot di động và thực tế tăng cường/thực tế ảo (AR/VR).