YOLO thích ứng hình ảnh để phát hiện vật thể trong điều kiện thời tiết bất lợi

Wenyu Liu1,2 *, Gaofeng Ren [3], Runsheng Yu [4], Shi Guo, [5] Jianke Zhu 1,2 †, Lei Zhang [3,5]

[1] Khoa Khoa học và Công nghệ Máy tính, Đại học Chiết Giang

[2] Đại học Alibaba-Chiết Giang Viện công nghệ biên giới liên kết

[3] Học viện DAMO, Tập đoàn Alibaba

[4] Đại học Khoa học và Công nghệ Hồng Kông

[5] Đại học Bách khoa HongKong

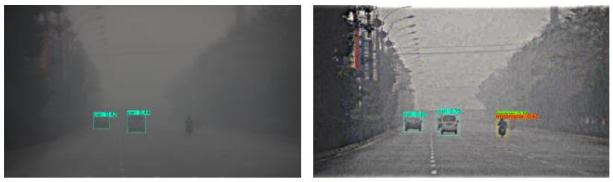
{liuwenyu.lwy, jkzhu}@zju.edu.cn, {gaof.ren, runninghengyu}@gmail.com, {csshiguo, cslzhang}@comp.polyu.edu.hk

trừu tượng

Mặc dù các phương pháp phát hiện đối tượng dựa trên học sâu có đạt được kết quả đầy hứa hẹn trên các bộ dữ liệu thông thường, nó là vẫn còn nhiều thách thức để xác định vị trí đối tượng từ hình ảnh chất lượng thấp bắt trong điều kiện thời tiết bất lợi. Các meth ods hiện có hoặc gặp khó khăn trong việc cân bằng các nhiệm vụ của hình ảnh nâng cao và phát hiện đối tượng, hoặc thường bỏ qua thông tin có lợi cho việc phát hiện. Để giảm bớt vấn đề này , chúng tôi đề xuất một YOLO thích ứng hình ảnh mới lạ (IA-YOLO) khung, trong đó mỗi hình ảnh có thể được cải tiến một cách thích ứng cho hiệu suất phát hiện tốt hơn. Cụ thể, một mô-đun xử lý tuổi tôi (DIP) có thể phân biệt được được trình bày để tính đến điều kiện thời tiết bất lợi cho máy dò YOLO, có các tham số được dự đoán bởi một công trình mạng nơ-ron tích tụ nhỏ (CNN-PP). Chúng ta cùng nhau học CNN-PP và YOLOv3 trong thời trang end-to-end, đảm bảo rằng CNN-PP có thể học một DIP thích hợp để nâng cao hình ảnh để phát hiện trong một cách thức giám sát yếu. Phương pháp tiếp cận IA-YOLO được đề xuất của chúng tôi có thể xử lý hình ảnh một cách thích ứng ở cả bình thường và bất lợi điều kiện thời tiết. Các kết quả thử nghiệm rất rõ ràng về sự lão hóa, chứng tỏ tính hiệu quả của phương pháp IA YOLO được đề xuất của chúng tôi trong cả tình huống sương mù và ánh sáng yếu. Các mã nguồn có thể được tìm thấy tại https://github.com/wenyyu/Image Adaptive-YOLO.

Giới thiệu

Các phương pháp dựa trên CNN đã trở nên phổ biến trong đối tượng detec tion (Ren et al. 2015; Redmon và Farhadi 2018). Họ không chỉ đạt được hiệu suất đầy hứa hẹn trên điểm chuẩn bộ dữ liệu (Deng và cộng sự 2009; Everedham và cộng sự 2010; Lin và cộng sự. 2014) nhưng cũng đã được triển khai trong các ứng dụng thực tế như lái xe tự động (Wang et al. 2019). Quá hạn đối với sự thay đổi miền trong hình ảnh đầu vào (Sindagi et al. 2020), các mô hình phát hiện đối tượng chung được đào tạo bởi các lứa tuổi chất lượng cao thường không đạt được kết quả khả quan trong điều kiện bất lợi điều kiện thời tiết (ví dụ: sương mù và ánh sáng tối). Narasimhan và Nayar (2002) và You et al. (2015) đề xuất rằng một tuổi tôi bị bắt trong điều kiện thời tiết bất lợi có thể bị phân hủy thành một hình ảnh rõ ràng và thông tin tương ứng về thời tiết cụ thể và chất lượng hình ảnh giảm sút trong thời tiết bất lợi



(a) YOLO II (Đường cơ sở) (b) IA-YOLO (Của chúng tôi)

Hình 1: Trong điều kiện sương mù trong thế giới thực, phương pháp của chúng tôi có thể đầu ra một cách thích ứng hình ảnh rõ ràng hơn với các cạnh xung quanh sắc nét hơn ranh giới của đối tượng, và do đó tạo ra kết quả phát hiện có độ chính xác cao hơn với ít trường hợp bị thiếu hơn.

chủ yếu là do sự tương tác giữa thời tiết cụ thể thông tin và đối tượng, dẫn đến khả năng phát hiện kém theo hình thức. Hình 1 cho thấy một ví dụ về phát hiện đối tượng trong điều kiện sương mù. Người ta có thể thấy rằng nếu hình ảnh có thể được nâng cao phù hợp với điều kiện thời tiết, hơn thế nữa thông tin tiềm ẩn về các đối tượng bị mờ ban đầu và các đối tượng xác định sai có thể được phục hồi.

Để giải quyết vấn đề khó khăn này, Huang, Le và Jaw (2020) đã sử dụng hai mạng con để cùng tìm hiểu tính năng nâng cao độ nghiêng trực quan và phát hiện đối tượng, trong đó tác động của sự suy giảm hình ảnh được giảm bớt bằng cách chia sẻ các lớp ngoại vi tính năng. Tuy nhiên, thật khó để điều chỉnh các thông số để cân bằng trọng lượng giữa phát hiện và phục hồi sau khi đào tạo. Một cách tiếp cận khác là giảm bớt ảnh hưởng của thông tin cụ thể về thời tiết bằng cách xử lý trước hình ảnh với các phương pháp hiện có như làm mờ ảnh (Hang et al. 2020; Liu và cộng sự. 2019) và nâng cao hình ảnh (Guo et al. 2020). Tuy nhiên, các mạng khôi phục hình ảnh phức tạp phải được đưa vào các phương pháp này, các phương pháp này cần được đào tạo riêng biệt với sự giám sát ở cấp độ pixel. Điều này yêu cầu phải làm theo cách thủ công gắn nhãn các hình ảnh để phục hồi. Vấn đề này cũng có thể được coi như một nhiệm vụ thích ứng miền không có giám sát (Chen et al. Năm 2018; Hnewa và Radha 2021). So với đào tạo máy dò có hình ảnh rõ ràng (hình ảnh nguồn), giả định rằng hình ảnh được chụp dưới thời tiết bất lợi (hình ảnh mục tiêu) có một sự thay đổi phân phối. Các phương pháp này chủ yếu áp dụng miền các nguyên tắc thích ứng và tập trung vào việc sắp xếp các tính năng của hai bản phân phối và thông tin tiềm ẩn có thể thu được trong quá trình khôi phục hình ảnh dựa trên thời tiết là

* Tác phẩm này được thực hiện khi tác giả đến thăm Alibaba với tư cách là một thực tập sinh nghiên cứu.

† Đồng tác giả

thường bị bỏ qua.

Để giải quyết những hạn chế trên, chúng tôi đề xuất một cách khéo léo phương pháp phát hiện đối tượng thích ứng hình ảnh, được gọi là IA-YOLO. Cụ thể, chúng tôi đề xuất một mô-đun nhập quy trình hình ảnh (DIP) hoàn toàn có thể phân biệt được, có siêu tham số có khả năng thích ứng được học bởi một công cụ dự đoán tham số nhỏ dựa trên CNN (CNN PP). CNN-PP dự đoán thích ứng độ lệch siêu tham số của DIP theo độ sáng, màu sắc, tông màu và thông tin thời tiết cụ thể của hình ảnh đầu vào. Sau khi xử lý bởi mô-đun DIP, sự can thiệp của thông tin thời tiết cụ thể trong hình ảnh có thể được khắc phục trong khi thông tin tiềm ẩn có thể được khôi phục. Chúng tôi trình bày một sơ đồ tối ưu hóa chung để tìm hiểu cách phát hiện đường trục DIP, CNN-PP và YOLOv3 mạng (Redmon và Farhadi 2018) trong một người đàn ông end-to-end ner. Để nâng cao hình ảnh để phát hiện, CNN-PP yếu được giám sát để tìm hiểu một DIP thích hợp thông qua các giới hạn chú thích hộp. Ngoài ra, chúng tôi sử dụng các hình ảnh trong cả điều kiện thời tiết bình thường và bất lợi để đào tạo mạng đề xuất. Bằng cách tận dụng lợi thế của mạng lưới CNN-PP , phương pháp tiếp cận IA-YOLO được đề xuất của chúng tôi có thể thích ứng đối phó với hình ảnh bị ảnh hưởng bởi các mức độ thời tiết khác nhau các điều kiện. Hình 1 cho thấy một ví dụ về kết quả phát hiện bằng phương pháp đề xuất của chúng tôi.

Những điểm nổi bật của công việc này là: 1) một khung kiểm tra thích ứng hình ảnh được đề xuất, đạt được hứa hẹn hiệu suất trong cả điều kiện thời tiết bình thường và bất lợi; 2) mô-đun xử lý hình ảnh có thể phân biệt trong hộp trắng là được đề xuất, có siêu tham số được dự đoán bởi một mạng dự báo tham số có giám sát; 3) khuyến khích kết quả thí nghiệm đạt được trên cả hai tấm thử nghiệm tổng hợp (VOC_Foggy và VOC_Dark) và tập dữ liệu trong thế giới thực (RTTS và ExDark), so sánh với các phương pháp trước đó.

Công việc liên quan

Phát hiện đối tượng

Là một nhiệm vụ cơ bản trong thị giác máy tính, việc dò tìm đối tượng đã nhận được sự quan tâm sâu sắc. Phát hiện đối tượng phương pháp có thể được chia thành hai loại (Zhao et al. 2019). Một loại là meth ods dựa trên đề xuất khu vực (Girshick và cộng sự 2014; Girshick 2015; Ren và cộng sự 2015), nơi đầu tiên tạo ra các vùng quan tâm (RoI) từ hình ảnh, và sau đó phân loại chúng bằng cách huấn luyện mạng nơ-ron. Nữa danh mục là các phương pháp tiếp cận dựa trên hồi quy một giai đoạn, chẳng hạn như Sê-ri YOLO (Redmon và cộng sự 2016; Redmon và Farhadi 2017, Năm 2018; Bochkovskiy, Wang và Liao 2020) và SSD (Liu et al. 2016), nơi tọa độ nhãn đối tượng và hộp giới hạn được dự đoán bởi một CNN. Trong bài báo này, chúng tôi sử dụng máy dò một giai đoạn cổ điển YOLOv3 (Redmon và Farhadi 2018) với tư cách là máy dò đường cơ sở và cải thiện hiệu suất của nó trong điều kiện bất lợi.

Thích ứng hình ảnh

Thích ứng hình ảnh được sử dụng rộng rãi trong việc nâng cao hình ảnh. Để nâng cao hình ảnh một cách thích hợp, một số meth ods truyền thống (Polesel, Ramponi và Mathews 2000; Yu và Bajaj Năm 2004; Wang và cộng sự. 2021) tính toán các thông số một cách thích ứng của sự biến đổi hình ảnh theo các đặc điểm tuổi im tương ứng. Ví dụ, Wang et al. (2021) đề xuất một chức năng điều chỉnh độ sáng điều chỉnh một cách thích ứng các thông số khởi động dựa trên sự phân bố ánh sáng đặc điểm của hình ảnh đầu vào.

Để đạt được nâng cao hình ảnh thích ứng, (Hu et al. 2018; Yu và cộng sự. Năm 2018; Zeng và cộng sự. 2020) đã thuê một CNN nhỏ để linh hoạt học các siêu tham số của phép biến đổi hình ảnh. Hu và cộng sự. (2018) đề xuất một khuôn khổ xử lý sau với tập hợp các bộ lọc có thể phân biệt, nơi học tập củng cố sâu (DRL) được sử dụng để tạo hoạt động và bộ lọc hình ảnh các thông số theo chất lượng của hiện tại được chỉnh sửa hình ảnh. Zeng và cộng sự. (2020) đã sử dụng một CNN nhỏ để tìm hiểu các LUT 3D thích ứng hình ảnh theo bối cảnh toàn cầu như độ sáng, màu sắc và tông màu.

Phát hiện đối tượng trong các điều kiện bất lợi

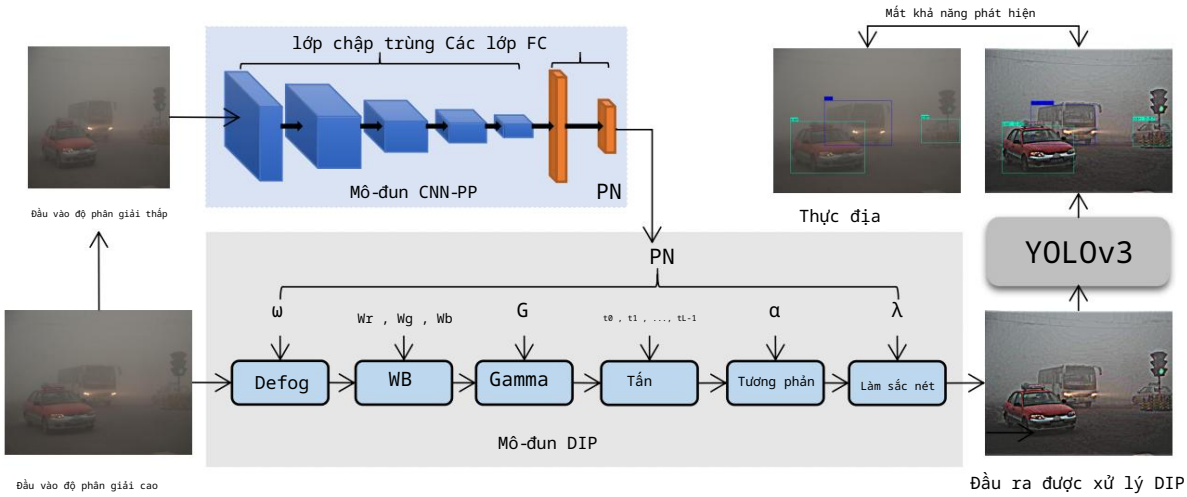
So với phát hiện đối tượng chung, ít nghiên cứu những nỗ lực đã được thực hiện để phát hiện đối tượng trong điều kiện bất lợi điều kiện thời tiết. Một cách tiếp cận đơn giản là xử lý trước hình ảnh bằng cách sử dụng hình ảnh hoặc chế độ khử ánh sáng cổ điển phương pháp nâng cao (Guo et al. 2020; He, Sun và Tang Năm 2009; Liu và cộng sự. Năm 2019; Hằng và cộng sự. Năm 2020; Qin và cộng sự. 2020), vốn được thiết kế để loại bỏ sương mù và nâng cao chất lượng tuổi thọ. Tuy nhiên, việc cải thiện chất lượng hình ảnh có thể chắc chắn không có lợi cho hiệu suất phát hiện. Một số phương pháp dựa trên cơ sở trước đây (Li et al. 2017; Huang, Le và Jaw 2020) cùng nhau thực hiện nâng cao và phát hiện hình ảnh để giảm bớt sự can thiệp của thông tin bất lợi về thời tiết cụ thể. Sindagi và cộng sự. (2020) đề xuất khung phát hiện đối tượng đối thủ miền để phát hiện trong điều kiện sương mù và mưa. Một vài phương pháp (Chen và cộng sự. Năm 2018; Zhang và cộng sự. Năm 2021; Hnewa và Radha 2021) tận dụng của miền thích ứng để giải quyết vấn đề này. Hnewa và Radha (2021) giả định rằng có sự thay đổi miền giữa các độ tuổi im được chụp trong điều kiện thời tiết bình thường và bất lợi. Họ đã thiết kế một YOLO thích ứng miền đa quy mô hỗ trợ thích ứng miền trong các lớp khác nhau tại tính năng giai đoạn chiết xuất.

Phương án đề xuất

Những hình ảnh được chụp trong điều kiện thời tiết bất lợi có tầm nhìn kém do sự giao thoa của thời tiết cụ thể trong quá trình hình thành, gây khó khăn trong việc phát hiện đối tượng. Để giải quyết thách thức này, chúng tôi đề xuất khung phát hiện thích ứng với hình ảnh hoạt động bằng cách xóa thông tin cụ thể về thời tiết và tiết lộ nhiều thông tin tiềm ẩn. Như minh họa trong Hình 2, toàn bộ đường ống bao gồm một bộ dự báo tham số dựa trên CNN (CNN PP), một mô-đun xử lý hình ảnh có thể phân biệt (DIP) và một mạng phát hiện. Đầu tiên, chúng tôi thay đổi kích thước hình ảnh đầu vào thành kích thước 256 × 256 và đưa nó vào CNN-PP để dự đoán DIP thông số. Sau đó, hình ảnh được lọc bởi mô-đun DIP sẽ được xử lý làm đầu vào cho máy dò YOLOv3. Chúng tôi giới thiệu một kết thúc từ đầu đến cuối chương trình đào tạo dữ liệu kết hợp với phát hiện mất mát để CNN-PP có thể học một DIP thích hợp để nâng cao hình ảnh để phát hiện đối tượng theo cách thức được giám sát yếu.

Mô-đun DIP

Như trong (Hu et al. 2018), việc thiết kế các bộ lọc hình ảnh nên tuân theo nguyên tắc phân biệt và phân giải-



Hình 2: Quy trình đào tạo end-to-end của khung IA-YOLO được đề xuất. Phương pháp của chúng tôi học một YOLO với một công cụ dự đoán tham số nhỏ dựa trên CNN (CNN-PP), sử dụng hình ảnh đầu vào được lấy mẫu xuống để dự đoán siêu tham số của các bộ lọc trong mô-đun DIP. Hình ảnh đầu vào có độ phân giải cao được bộ lọc của DIP xử lý để giúp YOLOv3 đạt được hiệu suất phát hiện cao. Bộ lọc Defog chỉ được sử dụng trong điều kiện sương mù.

sống độc lập. Đối với tối ưu hóa dựa trên gradient của CNN PP, các bộ lọc phải có thể phân biệt được để cho phép đào tạo mạng bằng cách nhân giống ngược. Vì CNN sẽ tiêu tốn rất nhiều tài nguyên máy tính để xử lý hình ảnh có độ phân giải cao (ví dụ: 4000 × 3000), trong bài báo này, chúng tôi tìm hiểu các thông số bộ lọc từ hình ảnh có độ phân giải thấp được lấy mẫu xuống có kích thước 256 × 256, và sau đó áp dụng cùng một bộ lọc cho hình ảnh có độ phân giải gốc. Do đó, các bộ lọc này cần độc lập với độ phân giải hình ảnh.

Mô-đun DIP được đề xuất của chúng tôi bao gồm sáu bộ lọc có thể phân biệt với các siêu thông số có thể điều chỉnh, bao gồm Defog, White Balance (WB), Gamma, Contrast, Tone và Sharpen. Như trong (Hu và cộng sự 2018), các phần tử opera màu và tông màu tiêu chuẩn, chẳng hạn như WB, Gamma, Độ tương phản và Tông màu, có thể được sử dụng dưới dạng các bộ lọc pixel khôn ngoan. Do đó, các bộ lọc được thiết kế của chúng tôi có thể được phân loại thành các bộ lọc Defog, Pixel-khôn ngoan và Sharpen. Trong số các bộ lọc này, bộ lọc Defog được thiết kế đặc biệt cho các cảnh có sương mù. Cac chi tiêt như sau.

Bộ lọc Pixel khôn ngoan. Bộ lọc pixel-khôn ngoan ánh xạ giá trị pixel đầu vào Pi = (ri , gi , bi) thành giá trị pixel đầu ra Po = (ro, go, bo), trong đó (r, g, b) đại diện cho các giá trị của ba màu các kênh màu đỏ, xanh lục và xanh lam, tương ứng.
Các chức năng ánh xạ của bốn bộ lọc pixel-khôn ngoan được liệt kê trong Bảng 1, trong đó cột thứ hai liệt kê các tham số cần được tối ưu hóa trong cách tiếp cận của chúng tôi. WB và Gamma là các phép nhân đơn giản và các phép biến đổi lũy thừa. Rõ ràng, các chức năng ánh xạ của chúng có thể phân biệt được đối với cả hình ảnh đầu vào và các tham số.

Các bộ lọc tương phản có thể phân biệt được thiết kế với một tham số đầu vào để đặt nội suy tuyến tính giữa hình ảnh gốc và hình ảnh được nâng cao hoàn toàn. Như thể hiện trong Bảng 1, định nghĩa của En (Pi) trong hàm ánh xạ như sau:

$$Lum(Pi) = 0{,}27ri + 0{,}67gi + 0{,}06bi \quad (1)$$

Lọc	Thông số	Chức năng lập bản đồ
WB	Wr, Wg, Wb: các hệ số	Po = (Wrri , Wggi , Wbbi)
Gamma	G: giá trị gamma	Po = P G
Độ tương phản	α: giá trị tương phản	Po = α · En (Pi) + (1 - α) · Pi
Tấn	ti : tone params	Po = (Ltr (ri), Ltg (gi), Ltb (bi))

Bảng 1: Các chức năng ánh xạ của bộ lọc pixel khôn ngoan.

$$EnLum(Pi) = \left(1 - \frac{1}{2}\cos(\pi \times (Lum(Pi)))\right) \quad (2)$$

$$En(Pi) = Pi \times \frac{EnLum(Pi)}{Lum(Pi)} \quad (3)$$

Như trong (Hu et al. 2018), chúng tôi thiết kế bộ lọc giai điệu như một hàm đơn điệu và tuyến tính từng đoạn. Chúng ta học tone fil ter với các tham số L, được biểu diễn là {t0, t1 , . . . , tL 1}. Các điểm của đường cong âm sắc được ký hiệu là ($k_i = i/L$, $i \in [0, L]$). Đối với hình ảnh đầu vào Pi, các tham số có thể phân biệt, điều này cho phép chức năng có thể phân biệt được đối với cả hình ảnh đầu vào và các tham số {t0, t1 , . . . , tL-1} như bên dưới

$$Po = \frac{1}{TL} \sum_{j=0}^{L-1} clip(L \cdot Pi - j, 0, 1)\, tk \quad (4)$$

Làm sắc nét bộ lọc. Làm sắc nét hình ảnh có thể làm nổi bật các chi tiết hình ảnh. Giống như kỹ thuật mặt nạ không mảnh (Polesel, Ram poni và Mathews 2000), quá trình làm sắc nét có thể được mô tả như sau:

$$F(x, \lambda) = I(x) + \lambda (I(x) - Gau(I(x))) \quad (5)$$

trong đó I (x) là hình ảnh đầu vào, Gau (I (x)) biểu thị bộ lọc Gaussian , và λ là một hệ số tỷ lệ dương. Độ sắc nét này có thể phân biệt được đối với cả x và λ. Lưu ý rằng mức độ làm sắc nét có thể được điều chỉnh để có hiệu suất phát hiện đối tượng tốt hơn bằng cách tối ưu hóa λ.

Bộ lọc Defog. Được thúc đẩy bởi phương pháp dark channel trước (He, Sun, và Tang 2009), chúng tôi thiết kế một tập tin defog với một tham số có thể học được. Dựa trên mô hình tán xạ khí quyển (McCartney 1976; Narasimhan và Nayar 2002), sự hình thành của một hình ảnh mờ có thể được xây dựng như sau:

$$I(x) = J(x)\,t(x) + A\,(1 - t(x)) \quad (6)$$

trong đó I (x) là hình ảnh sương mù và J (x) đại diện cho sự rạng rỡ của cảnh (hình ảnh sạch). A là ánh sáng khí quyển toàn cầu và t (x) là ánh sáng truyền môi trường, được định nghĩa là: $t(x) = e^{-\beta d(x)}$ (7) trong đó β (x) đại diện cho hệ số tán xạ của quả cầu atmo và d (x) là chiều sâu cảnh.

Để khôi phục ảnh sạch J (x), điều quan trọng là phải thu được ánh sáng khí quyển A và ánh xạ truyền t (x). Để đạt được mục đích này, trước tiên chúng tôi tính toán bản đồ kênh tối của hình ảnh sương mù I (x) và chọn 1000 pixel sáng nhất hàng đầu. Sau đó, A được ước tính bằng cách lấy trung bình 1000 pixel đó của vị trí tương ứng của hình ảnh sương mù I (x).

Theo Eq. (6), người ta có thể suy ra một solu gần đúng tion of t (x) như sau (He, Sun, and Tang 2009)

$$t(x) = 1 - \min_{C} \min_{y \in \Omega(x)} \frac{I^{dark}_{C}(y)}{A_C} \quad (số\ 8)$$

Chúng tôi giới thiệu thêm một tham số ω để kiểm soát mức độ làm mờ như sau:

$$t(x, \omega) = 1 - \omega \min_{C} \min_{y \in \Omega(x)} \frac{I^{dark}_{C}(y)}{A_C} \quad (9)$$

Vì hoạt động trên có thể phân biệt được, chúng ta có thể tối ưu hóa ω thông qua lan truyền ngược để làm cho bộ lọc defog linh hoạt hơn để phát hiện hình ảnh có sương mù.

Mô-đun CNN-PP Trong

đường ống xử lý tín hiệu hình ảnh camera (ISP), một số bộ lọc quảng cáo vừa ý thường được sử dụng để nâng cao hình ảnh, có các siêu thông số được các kỹ sư có kinh nghiệm điều chỉnh thủ công thông qua kiểm tra trực quan (Mosleh và cộng sự 2020). Nói chung, quá trình điều chỉnh như vậy là rất khó khăn và phải mất nhiều thời gian để tìm các thông số phù hợp cho nhiều cảnh. Để giải quyết hạn chế này, chúng tôi đề xuất sử dụng một CNN nhỏ làm công cụ dự đoán tham số để ước tính các siêu tham số, điều này rất hiệu quả.

Lấy cảnh sương mù làm ví dụ, mục đích của CNN-PP là dự đoán các thông số của DIP bằng cách hiểu nội dung toàn cầu của hình ảnh, chẳng hạn như độ sáng, màu sắc và tông màu, cũng như mức độ sương mù. Do đó, hình ảnh được lấy mẫu xuống là đủ để ước tính những thông tin này, điều này có thể giúp tiết kiệm đáng kể chi phí tính toán. Với hình ảnh đầu vào của bất kỳ độ phân giải nào, chúng tôi chỉ cần sử dụng phép nội suy song tuyến để giảm mẫu xuống độ phân giải 256 × 256 . Như thể hiện trong Hình 2, mạng CNN-PP bao gồm năm khối tích hợp và hai lớp được kết nối đầy đủ. Mỗi khối chập bao gồm một lớp chập 3 × 3 với bước 2 và một Rơle bình thường. Lớp cuối cùng được kết nối đầy đủ xuất ra siêu tham số cho mô-đun DIP. Kênh đầu ra của năm lớp hội tụ này lần lượt là 16, 32, 32, 32 và 32. Mô hình CNN-PP chỉ chứa tham số 165K khi tổng số tham số là 15.

Mô-đun mạng phát hiện

Trong bài báo này, chúng tôi chọn máy dò một giai đoạn YOLOv3 làm mạng phát hiện, được sử dụng rộng rãi trong các thiết bị thực tế , bao gồm chỉnh sửa hình ảnh, giám sát an ninh, phát hiện đám đông và lái xe tự động (Zhang et al. 2021). Tương tự như phiên bản trước, YOLOv3 thiết kế darknet-53 bao gồm các lớp màu phức tạp 3 × 3 và 1 × 1 liên tiếp dựa trên ý tưởng của Resnet (He et al. 2016). Nó thực hiện đào tạo đa tỷ lệ bằng cách đưa ra dự đoán trên bản đồ đối tượng nhiều tỷ lệ, để cải thiện hơn nữa độ chính xác của việc phát hiện, đặc biệt là đối với các vật thể nhỏ. Chúng tôi áp dụng cùng một kiến trúc mạng và các chức năng mất mát như YOLOv3 ban đầu (Redmon và Farhadi 2018).

Đào tạo dữ liệu kết hợp Để

đạt được hiệu suất phát hiện lý tưởng trong cả điều kiện thời tiết bất lợi và bất lợi, chúng tôi áp dụng một chương trình đào tạo dữ liệu kết hợp cho IA-YOLO. Thuật toán 1 tóm tắt quá trình đào tạo của phương pháp đề xuất của chúng tôi. Mỗi hình ảnh có xác suất 2/3 được thêm ngẫu nhiên với một số loại sương mù hoặc được chuyển thành hình ảnh thiếu sáng trước khi được đưa vào mạng để huấn luyện. Với cả dữ liệu đào tạo chất lượng thấp theo chủ đề thông thường và tổng hợp, toàn bộ đường ống được đào tạo từ đầu đến cuối với tổn thất phát hiện YOLOv3, điều này đảm bảo tất cả các mô-đun trong IA-YOLO có thể thích ứng với nhau. Do đó, mô-đun CNN-PP được giám sát yếu bởi sự mất mát phát hiện mà không gắn nhãn hình ảnh chân thực mặt đất theo cách thủ công. Chế độ huấn luyện dữ liệu kết hợp đảm bảo rằng IA-YOLO có thể xử lý hình ảnh một cách thích ứng theo nội dung của từng hình ảnh, để đạt được hiệu suất phát hiện cao.

Thí nghiệm

Chúng tôi đánh giá hiệu quả của phương pháp của chúng tôi trong các tình huống sương mù và ánh sáng yếu. Tổ hợp bộ lọc là [Defog, White Balance (WB), Gamma, Contrast, Tone, Shapen], trong khi bộ lọc Defog chỉ được sử dụng trong điều kiện sương mù.

Chi tiết triển khai Chúng

tôi áp dụng giao thức đào tạo của (Redmon và Farhadi 2018) trong cách tiếp cận IA-YOLO được đề xuất của chúng tôi. Mạng xương sống cho tất cả các thí nghiệm là Darknet-53. Trong quá trình đào tạo, Chúng tôi đã chạy thay đổi kích thước hình ảnh thành (32N × 32N), trong đó N [9, 19]. Hơn nữa, các phương pháp tăng dữ liệu như lật, cắt và biến đổi hình ảnh được áp dụng để mở rộng tập dữ liệu đào tạo. Mô hình IA-YOLO của chúng tôi được đào tạo bởi trình tối ưu hóa Adam (Kingma và Ba 2014) với 80 kỷ nguyên. Tỷ lệ học bắt đầu là 10 4 và kích thước lô là 6. IA YOLO dự đoán các hộp giới hạn ở ba tỷ lệ khác nhau và ba điểm neo ở mỗi tỷ lệ. Chúng tôi sử dụng Tensorflow cho các thử nghiệm của mình và chạy nó trên GPU Tesla V100.

Thử nghiệm trên tập dữ liệu hình ảnh

sương mù Có rất ít tập dữ liệu có sẵn công khai để phát hiện ob ject trong điều kiện thời tiết bất lợi và số lượng dữ liệu thường nhỏ để đào tạo một máy dò dò tìm dựa trên CNN ổn định. Để tạo điều kiện so sánh công bằng, chúng tôi xây dựng dựa trên tập dữ liệu VOC cổ điển (Everedham và cộng sự 2010) một tập dữ liệu VOC_Foggy theo mô hình tán xạ khí quyển (Narasimhan

Thuật toán 1: Quy trình đào tạo YOLO thích ứng với hình ảnh

Khởi tạo mạng CNN-PP P_θ và mạng YOLOv3 làm việc D_β với trọng số ngẫu nhiên θ và β.
Đặt giai đoạn đào tạo: num_epochs = 80, batch_size = 6.

Chuẩn bị tập dữ liệu bình thường VOC_norm_trainval. cho tôi trong num_epochs làm lặp lại Chụp một loạt ảnh M từ VOC_norm_trainval; đối với j trong batch_size thực hiện nếu random.randint (0, 2)> 0 thì Tạo hình ảnh sương mù M (j) (Phương trình (10, 11)), trong đó A = 0,5, k = random.randint (0, 9) , β = 0,01 k + 0,05 // đối với điều kiện sương mù Tạo ảnh thiếu sáng M (j) bằng f (x) = trong đó γ = random.uniform (1,5, 5) // đối với điều kiện ánh sáng yếu end if end for Compute DIP params by PN = P Thực hiện xử lý bộ lọc DIP bởi image_batch = DIP (image_batch, PN); Gửi ảnh_bố tới mạng YOLOv3 D_β ; Cập nhật mạng CNN-PP P và mạng YOLOv3 D_β theo tổn thất phát hiện YOLOv3.

$_{x,y}$ $_\theta$ (image_batch);

$_\theta$

cho đến khi tất cả các hình ảnh đã được đưa vào các mô hình đào tạo thì kết thúc cho

Hình ảnh tập dữ liệu	ps	bc	xe hơi	tổng mc bus
V_n_tv	8111	13256	1064	3267 822 1052 19561
V_n_ts	2734	4528	337	1201 213 325 6604
RTTS	4322	7950	534	18413 1838 862 29577

Bảng 2: Thống kê các bộ dữ liệu đã sử dụng, bao gồm V_n_tv (VOC_norm_trainval), V_n_ts (VOC_norm_test) và RTTS. Các lớp học là ps (người), bc (xe đạp), ô tô, xe buýt và mc (xe máy).

và Nayar 2002). Hơn nữa, RTTS (Li et al. 2018) là một tập dữ liệu tương đối toàn diện trong thế giới thực có sẵn trong điều kiện sương mù , có 4.322 hình ảnh mờ tự nhiên với năm lớp đối tượng được chú thích, cụ thể là người, xe đạp, ô tô, xe buýt và xe máy. Để tạo tập dữ liệu đào tạo của chúng tôi, chúng tôi chọn dữ liệu có chứa năm danh mục này để thêm sương mù.

Đối với VOC2007_trainval và VOC2012_trainval, chúng tôi lọc các hình ảnh bao gồm năm lớp đối tượng ở trên để xây dựng VOC_norm_trainval. VOC_norm_test được chọn từ VOC2007_test theo cách tương tự. Chúng tôi cũng đánh giá phương pháp của chúng tôi trên RTTS. Thống kê của các bộ dữ liệu được tổng hợp thành Bảng 2.

Để tránh chi phí tính toán khi tạo hình ảnh sương mù trong quá trình đào tạo, chúng tôi xây dựng tập dữ liệu VOC_Foggy ngoại tuyến. Theo Eqs. (6, 7), ảnh sương mù I (x) thu được bằng cách

$$I(x) = J(x)\, e^{d(x)\beta} + A\,(1 - e^{d(x)\beta}) \qquad (10)$$

Phương pháp	V_n_ts	V_F_t	RTTS
Đường cơ sở	YOLOv3 I VOC_norm 70.10 31,05 28,82 YOLOv3 II Dữ liệu kết hợp 64,13 63,40 30,80 57,38 30,20 MSBDN VOC_norm GridDehaze VOC_norm 58,23 31,42 DAYOLO Dữ liệu lai 56,51 55,11 29,93		
Defog	Dữ liệu lai 53,21 73,23 IA- YOLO		
DA			
DSNet đa tác vụ			
Của chúng tôi			

Bảng 3: So sánh hiệu suất trên ảnh sương mù. "DA" có nghĩa là Thích ứng miền. Ba cột bên phải liệt kê mAP trên ba tập dữ liệu thử nghiệm, bao gồm V_n_ts (VOC_norm_test), V_F_t (VOC_Foggy_test) và RTTS.

d (x) được định nghĩa như sau:

$$d(x) = 0{,}04\,\rho + p \max_{(row,\,col)\ \text{trong}} \qquad (11)$$

đó ρ là khoảng cách Euclide từ pixel hiện tại đến pixel trung tâm, row và col lần lượt là số hàng và cột của hình ảnh. Bằng cách đặt A = 0,5 và β = 0,01 i + 0,05, trong đó i là số nguyên từ 0 đến 9, có thể thêm mười mức sương mù khác nhau vào mỗi hình ảnh. Dựa trên tập dữ liệu VOC_norm_trainval, chúng tôi tạo ra tập dữ liệu VOC_Foggy_trainval lớn hơn gấp mười lần so với mặt phẳng . Để có được tập dữ liệu VOC_Foggy_test, mọi hình ảnh trong VOC_norm_test được xử lý ngẫu nhiên với sương mù.

Kết quả thử nghiệm Để chứng minh tính hiệu quả của IA-YOLO, chúng tôi so sánh phương pháp của mình với YOLOv3, Defog + Detect, điều chỉnh miền (Hnewa và Radha 2021) và học đa tác vụ (Huang, Le và Jaw 2020) trong ba thử nghiệm bộ dữ liệu. Đối với Defog + Detect, chúng tôi sử dụng phương pháp khử mờ làm bước tiền xử lý và sử dụng YOLOv3 được đào tạo trên VOC_norm để phát hiện. Chúng tôi chọn MSBDN (Hang et al. 2020) và GridDehaze (Liu et al. 2019) làm phương pháp xử lý mỗi, là những phương pháp dehazing dựa trên CNN phổ biến. Đối với phương pháp điều chỉnh miền, chúng tôi triển khai công cụ dò tìm miền thích ứng đa quy mô một giai đoạn DAY OLO (Hnewa và Radha 2021) với nhiều đường dẫn điều chỉnh miền và các bộ phân loại miền tương ứng ở các quy mô khác nhau của YOLOv3. Đối với các thuật toán học tập đa tác vụ, chúng tôi chọn DSNet (Huang, Le và Jaw 2020) cùng học cách phát hiện và tắt đèn trong điều kiện thời Chúng tôi tái tạo mô hình khôi phục và mạng con phát hiện của nó bằng cách chia sẻ năm lớp chập đầu tiên của Yolov3 và cùng huấn luyện hai mạng với dữ liệu kết hợp.

Bảng 3 so sánh độ chính xác trung bình trung bình (mAP) giảm xuống (Everedham và cộng sự 2012) giữa IA-YOLO và các thuật toán cạnh tranh khác ở cả điều kiện bình thường và mơ hồ. Cột thứ hai liệt kê dữ liệu huấn luyện cho từng phương pháp, trong đó "Dữ liệu kết hợp" có nghĩa là sơ đồ nhập huấn luyện dữ liệu kết hợp được sử dụng trong IA-YOLO được đề xuất của chúng tôi. So với đường cơ sở (YOLO I), tất cả các phương pháp đều có những cải tiến trên cả bộ dữ liệu kiểm tra thời tiết sương mù tổng hợp và từ thực, trong khi chỉ có IA-YOLO của chúng tôi không bị suy giảm trong trường hợp bình thường. Điều này là do các phương pháp trước đây chủ yếu được thiết kế để xử lý phát hiện vật thể trong điều kiện thời tiết sương mù trong khi hy sinh hiệu suất của chúng trên hình ảnh thời tiết bình thường. Đối với phương pháp IA-YOLO được đề xuất của chúng tôi, các mô-đun CNN-PP và DIP có thể xử lý một cách thích

Hình 3: Kết quả phát hiện YOLOv3 II (cột 1, 3 và 5) và IA-YOLO (cột 2, 4 và 6) của chúng tôi trên tổng hợp Hình ảnh VOC_Foggy_test (hàng trên) và hình ảnh sương mù RTTS trong thế giới thực (hàng dưới). Phương pháp được đề xuất học cách làm mờ sương mù và làm sắc nét cạnh hình ảnh, có hiệu suất phát hiện tốt hơn với ít phát hiện sai và bỏ sót hơn.

Hình 4: Kết quả phát hiện của YOLOv3 II (cột bên trái) và IA-YOLO của chúng tôi (cột bên phải) trên VOC_Dark_test tổng hợp hình ảnh (hàng trên cùng) và hình ảnh ánh sáng yếu ExDark trong thế giới thực (hàng dưới cùng). Phương pháp được đề xuất học để nâng cao hình ảnh tương phản với nhiều chi tiết hơn.

hình ảnh với các mức độ sương mù khác nhau để phát hiện đối tượng. Kết quả là, phương pháp tiếp cận IA-YOLO được đề xuất của chúng tôi vượt trội hơn tất cả các phương pháp cạnh tranh trên ba tập dữ liệu thử nghiệm bởi một lượng lớn lợi nhuận, thể hiện hiệu quả của nó trong việc phát hiện đối tượng trong điều kiện thời tiết bất lợi.

Hình 3 cho thấy một số ví dụ trực quan về IA-YOLO của chúng tôi phương pháp và đường cơ sở YOLOv3 II. Mặc dù đối với một số trường hợp, mô-đun DIP thích ứng của chúng tôi tạo ra một số tiếng ồn cho nhận thức thị giác, nó tăng đáng kể độ dốc của tuổi tôi cục bộ dựa trên ngữ nghĩa hình ảnh và dẫn đến tốt hơn hiệu suất phát hiện.

Thử nghiệm hình ảnh trong điều kiện ánh sáng yếu

Bộ dữ liệu PSCAL VOC (Everedham et al. 2010) và Bộ dữ liệu phát hiện ánh sáng yếu tương đối toàn diện Ex Dark (Loh và Chan 2019) cả hai đều chứa mười danh mục các đối tượng: Xe đạp, Thuyền, Chai, Xe buýt, Xe hơi, Mèo, Ghế, Con chó, Xe máy, Con người (Người). Từ VOC2007_trainval và VOC2012_trainval, chúng tôi đã lọc các hình ảnh bao gồm bất kỳ lớp nào trong số mười lớp đối tượng ở trên để xây dựng VOC_norm_trainval. VOC_norm_test được chọn từ

Dữ liệu tàu	phương thức	V_n_ts	V_D_t	E_t
Đường cơ sở	YOLOv3 I	VOC_norm	69,13	45,92 36,42
	Dữ liệu kết hợp	YOLOv3 II	65,33	52,28 37,03
Tăng cường	ZeroDCE	VOC_norm	33.57	34.41 /
DA	DAYOLO	Dữ liệu kết hợp	41,68	21,53 18,15
Dữ liệu kết hợp	DSNet	đa tác vụ	64,08	43,75 36,97
Dữ liệu kết hợp	IA-YOLO	của chúng tôi	70,02	59,40 40,37

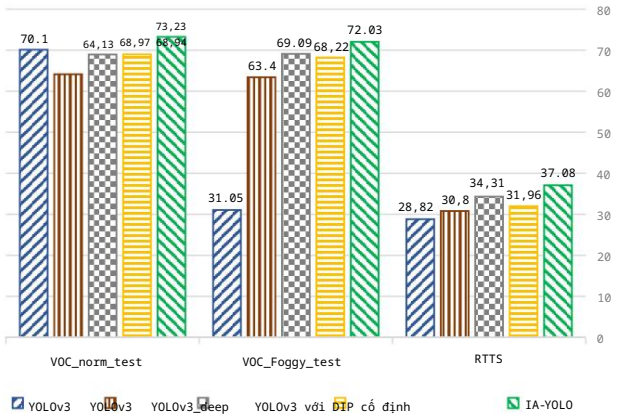
Bảng 4: So sánh hiệu suất trên ảnh thiếu sáng. Các ba cột bên phải liệt kê mAP trên ba tập dữ liệu thử nghiệm, bao gồm V_n_ts (VOC_norm_test), V_D_t (VOC_Dark_test) và E_t (Exdark_test).

VOC2007_test theo cách tương tự. Tổng số hình ảnh trong VOC_norm_trainval, VOC2007_test và ExDark_test là 12334, 3760 và 2563, tương ứng.

Chúng tôi tổng hợp tập dữ liệu VOC_dark ánh sáng yếu dựa trên VOC_norm thông qua phép biến đổi $f(x) = x^{\gamma}$, ở đâu giá trị của γ được lấy mẫu ngẫu nhiên từ một phân phối đồng nhất trong phạm vi [1,5, 5] và x biểu thị pixel đầu vào cường độ.

Kết quả thử nghiệm Chúng tôi so sánh phương pháp IA YOLO đã trình bày của chúng tôi với phương pháp YOLOv3 cơ bản, Nâng cao + Phát hiện, DAYOLO và DSNet trên ba bộ dữ liệu thử nghiệm. Vì Nâng cao + Phát hiện, chúng tôi sử dụng phương pháp nâng cao hình ảnh gần đây Zero-DCE (Guo et al. 2020) để xử lý trước hình ảnh ánh sáng yếu và sử dụng YOLOv3 được đào tạo trên VOC_norm để phát hiện. Các cài đặt thử nghiệm còn lại được giữ nguyên giống như những hình ảnh trên hình ảnh sương mù. Bảng 4 cho thấy mAP kết quả. Có thể thấy rằng phương pháp của chúng tôi mang lại kết quả tốt nhất. IA-YOLO cải thiện YOLO I đường cơ sở thêm 0,89, 13,48 và 3,95 phần trăm trên VOC_norm_test, VOC_Dark_test và ExDark_test, tương ứng, và nó cải thiện đường cơ sở của YOLO II thêm 4,69, 7,12 và 3,34 phần trăm trên các bộ thử nghiệm đó. Đây chứng minh rằng phương pháp tiếp cận IA-YOLO được đề xuất của chúng tôi cũng là hiệu quả trong điều kiện ánh sáng yếu.

Hình 4 cho thấy các so sánh định tính giữa IA YOLO và YOLOv3 II đường cơ sở. Có thể quan sát thấy rằng mô-đun DIP được đề xuất của chúng tôi có thể tăng độ tương phản của hình ảnh đầu vào và tiết lộ chi tiết hình ảnh, rất cần thiết để phát hiện đối tượng.



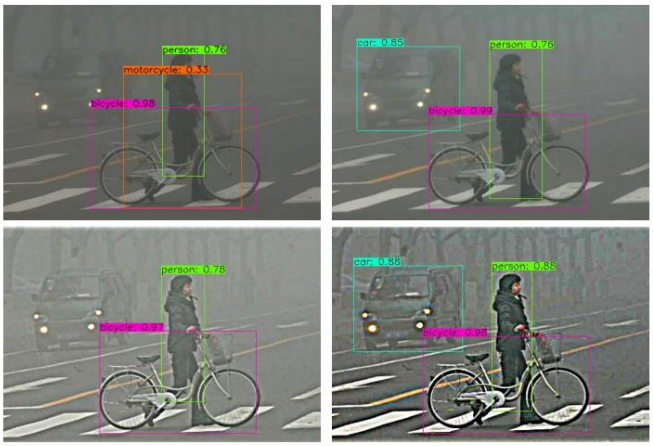
Hình 5: So sánh hiệu suất của các cài đặt khác nhau trong điều kiện sương mù.

Nghiên cứu cắt bỏ

Để kiểm tra tính hiệu quả của từng mô-đun trong đề xuất của chúng tôi khuôn khổ, chúng tôi tiến hành các thử nghiệm cắt bỏ trên các cài đặt, bao gồm chương trình đào tạo dữ liệu kết hợp, DIP và điều chỉnh theo độ tuổi. Chúng tôi cũng đánh giá việc lựa chọn đề xuất các bộ lọc có thể phân biệt trên ba tập dữ liệu thử nghiệm.

Kết quả của các thí nghiệm đã tiến hành được mô tả trong Hình 5. Ngoại trừ YOLO tôi đã đào tạo với VOC_norm, phần còn lại trong số các thử nghiệm sử dụng cùng một khóa đào tạo dữ liệu kết hợp và cài đặt thử nghiệm. Có thể thấy rằng đào tạo dữ liệu kết hợp, Tất cả các phương pháp xử lý trước bộ lọc DIP và các phương pháp thích ứng hình ảnh đều có thể cải thiện hiệu suất phát hiện trên cả VOC_Foggy_test và RTTS so với YOLO I. IA-YOLO đạt được kết quả tốt nhất bằng cách sử dụng cả ba mô-đun. YOLOv3 với DIP cố định có nghĩa là các siêu tham số của bộ lọc là một tập hợp các giá trị cố định, tất cả đều nằm trong phạm vi hợp lý phạm vi. YOLOv3_deep II là phiên bản sâu hơn của YOLO II bằng cách thêm tám lớp tích chập với các tham số học trên 411K. Như thể hiện trong Hình 5, IA-YOLO được đề xuất của chúng tôi phương pháp tiếp cận hoạt động tốt hơn YOLOv3_deep II chỉ với 165K tham số bổ sung trong CNN-PP. Những người đàn ông xứng đáng cho rằng chỉ có mô-đun học tập thích ứng mới cải thiện hiệu suất trên VOC_norm_test so với YOLO Tôi trong điều kiện thời tiết bình thường, trong khi cả YOLOv3 II và YOLOv3 với DIP cố định thu được kết quả kém hơn. Đây chứng minh rằng IA-YOLO có thể xử lý một cách thích ứng cả hình ảnh bình thường và hình ảnh có sương mù, có lợi cho nhiệm vụ phát hiện xuống dòng.

Như thể hiện trong Bảng 5, chúng tôi tiến hành đánh giá mAP định lượng về việc lựa chọn bộ lọc bằng cách sử dụng ba bộ dữ liệu thử nghiệm. Qua kết hợp ba bộ lọc, Model D thu được bộ lọc tốt nhất kết quả, chứng minh hiệu quả của các bộ lọc này. Hình 6 cho thấy các so sánh trực quan của một số mô hình trong Bảng 5. So với Model C, làm sáng và sắc nét hình ảnh và Mô hình B làm mờ hình ảnh, hình ảnh được xử lý bởi Model D không chỉ sáng hơn và sắc nét hơn mà cũng rõ ràng hơn, làm cho các đối tượng sương mù mờ hơn nhiều . Hơn nữa, chúng tôi cung cấp một số ví dụ về cách CNN-PP dự đoán các thông số của mô-đun DIP trong phần bổ sung



Hình 6: Kết quả phát hiện trên RTTS. Từ trái sang phải, trên cùng xuống dưới cùng: YOLO II, Model B, Model C và Model D.

Người mẫu	Defog Pixel_wise Sharpen Filter Bộ lọc Bộ lọc	V_n_ts V_F_t RTTS
Một	71,47 70,25 34,88	
B	71,43 70,14 34,83	
C		70,51 70,09 34,95
D		73,23 72,03 37,08

Bảng 5: Phân tích cắt bỏ trên các bộ lọc trong mô-đun DIP.

vật chất. Vui lòng tham khảo tệp bổ sung để biết thêm chi tiết.

Phân tích hiệu quả

Trong khuôn khổ IA-YOLO của chúng tôi, chúng tôi giới thiệu một mô-đun học tập CNN PP nhỏ với 165K tham số có thể huấn luyện vào YOLOv3. IA-YOLO mất 44 mili giây để phát hiện 544 × 544 × 3 hình ảnh độ phân giải trên một GPU Tesla V100 duy nhất. Nó chỉ có giá bổ sung 13 mili giây so với đường cơ sở YOLOv3, trong khi đó là 7 mili giây và nhanh hơn 50 ms so với GridDehaze-YOLOv3 và MSBDN YOLOv3, tương ứng. Tóm lại, IA-YOLO chỉ bổ sung 165K thông số có thể đào tạo trong khi đạt được hiệu suất tốt hơn nhiều trên tất cả các bộ dữ liệu thử nghiệm với thời gian chạy tương đương.

Sự kết luận

Chúng tôi đã đề xuất một cách tiếp cận IA-YOLO mới để cải thiện đối tượng phát hiện trong điều kiện thời tiết bất lợi, nơi mỗi đầu vào hình ảnh đã được cải tiến một cách thích ứng để có được khả năng phát hiện tốt hơn màn biểu diễn. Một mô-đun xử lý hình ảnh hoàn toàn khác biệt được phát triển để khôi phục nội dung tiềm ẩn bằng cách xóa thông tin cụ thể về thời tiết cho máy dò YOLO, có thông số đo hy được dự đoán bằng một dây thần kinh xoắn nhỏ mạng. Hơn nữa, toàn bộ khuôn khổ đã được đào tạo trong một thời trang end-to-end, nơi mạng lưới dự đoán tham số được giám sát yếu để học một mô-đun DIP thích hợp thông qua phát hiện mất mát. Bằng cách tận dụng lợi thế của lai đào tạo và mạng dự đoán tham số, đề xuất của chúng tôi phương pháp tiếp cận đã có thể xử lý một cách thích ứng các điều bình thường và bất lợi điều kiện thời tiết. Kết quả thử nghiệm cho thấy phương pháp thực hiện tốt hơn nhiều so với các phương pháp trước đó trong cả tình huống sương mù và thiếu sáng.

Lời cảm ơn Công trình

này được hỗ trợ bởi Tổ chức Khoa học Tự nhiên Quốc gia của Trung Quốc theo Grants (61831015) và Hong Kong RGC RIF Grant (R5001-18). Công việc này cũng được hỗ trợ bởi sự tài trợ của "Nhóm đổi mới hàng đầu của tỉnh Chiết Giang " (Grant NO. 2018R01017).

Người giới thiệu

Bochkovskiy, A. .; Wang, C.-Y .; và Liao, H.-YM 2020. Yolov4: Tốc độ và độ chính xác phát hiện đối tượng tối ưu. arXiv: 2004.10934.

Chen, Y. Li, W .; Sakaridis, C.; Đại, D.; và Van Gool, L. 2018. Tên miền r-cnn thích ứng nhanh hơn để phát hiện đối tượng trong tự nhiên. Trong Kỷ yếu của IEEE / CVF Conference Computer Vision Pattern Recognition (CVPR), 3339–3348.

Đặng, J.; Đồng, W .; Socher, R .; Li, L.-J.; Li, K.; và Fei-Fei, L. 2009. Imagenet: Cơ sở dữ liệu hình ảnh phân cấp quy mô lớn. Trong Kỷ yếu của IEEE / CVF Conference Computer Vision Pattern Recognition (CVPR), 248–255. IEEE.

Evingham, M.; Van Gool, L.; Williams, CK; Winn, J .; và Zisserman, A. 2010. Thử thách các lớp đối tượng trực quan pascal (voc). Tạp chí Quốc tế về Thị giác Máy tính, 88 (2): 303–338.

Evingham, M.; Van Gool, L.; Williams, CKI; Winn, J .; và Zisserman, A. 2012. Kết quả PASCAL Visual Object Classes Challenge 2012 (VOC2012). http: //www.pascal network.org/challenges/VOC/voc2012/workshop/index.html.

Girshick, R. 2015. Nhanh chóng r-cnn. Trong Kỷ yếu của Hội nghị Quốc tế IEEE về Thị giác Máy tính (ICCV), 1440– 1448.

Girshick, R .; Donahue, J .; Darrell, T.; và Malik, J. 2014. Hệ thống phân cấp tính năng phong phú để phát hiện đối tượng chính xác và phân đoạn ngữ nghĩa. Trong Kỷ yếu của IEEE / CVF Confer ence Computer Vision Pattern Recognition (CVPR), 580–587.

Guo, CG; Li, C.; Guo, J .; Loy, CC; Hou, J .; Kwong, S.; và Cong, R. 2020. Ước tính đường cong sâu không tham chiếu để nâng cao hình ảnh trong điều kiện ánh sáng yếu. Trong Kỷ yếu của Hội nghị IEEE / CVF Nhận dạng Mẫu Thị giác Máy tính (CVPR), 1780–1789.

Hằng, D.; Kim Sơn, P.; Zhe, H.; Xiang, L.; Xinyi, Z .; Fei, W .; và Ming-Hsuan, Y. 2020. Mạng lưới khử mùi tăng cường đa quy mô với tính năng kết hợp dày đặc. Trong Kỷ yếu của Hội nghị IEEE / CVF Nhận dạng Mẫu Thị giác Máy tính (CVPR).

Anh ấy, K .; CN, J .; và Tang, X. 2009. Hình ảnh duy nhất tái hiện sương mù bằng cách sử dụng kênh tối trước đó. Trong Kỷ yếu của Hội nghị IEEE / CVF Nhận dạng Mẫu Thị giác Máy tính (CVPR).

Anh ấy, K .; Zhang, X. .; Ren, S.; và Sun, J. 2016. Học tập sâu để nhận dạng hình ảnh. Trong Kỷ yếu của IEEE / CVF Conference Computer Vision Pattern Recognition (CVPR), 770–778.

Hnewa, M.; và Radha, H. 2021. Đa tỷ lệ Thực hiện YOLO thích ứng chính để phát hiện đối tượng trên nhiều miền. arXiv: 2106.01483.

Hu, Y .; Anh ấy, H.; Xu, C.; Vương, B.; và Lin, S. 2018. Expo chắc chắn: Khung xử lý hậu kỳ ảnh hộp trắng. Giao dịch ACM trên Đồ họa (TOG), 37 (2): 26.

Huang, S.-C.; Lê, T.-H.; và Jaw, D.-W. 2020. DSNet: Học ngữ nghĩa chung để phát hiện đối tượng trong điều kiện thời tiết khắc nghiệt . Giao dịch IEEE trên Phân tích mẫu và Ma chine Intelligence.

Kingma, DP; và Ba, J. 2014. Adam: Một phương pháp tối ưu hóa stochas tic. arXiv: 1412.6980.

Li, B.; Peng, X.; Wang, Z .; Xu, J .; and Feng, D. 2017. Aod net: Mạng khử mùi tất cả trong một. Trong Kỷ yếu của Hội nghị Quốc tế IEEE về Thị giác Máy tính (ICCV), 4770–4778.

Li, B.; Ren, W .; Fu, D.; Tao, D.; Feng, D.; Zeng, W .; và Wang, Z. 2018. Đánh giá điểm chuẩn cho hình ảnh đơn lẻ và hơn thế nữa. Giao dịch IEEE về Xử lý hình ảnh, 28 (1): 492–505.

Lin, T.-Y .; Maire, M.; Belongie, S.; Hays, J .; Perona, P.; Ramanan, D.; Dollár, P.; và Zitnick, CL 2014. Microsoft coco: Đối tượng chung trong ngữ cảnh. Trong Hội nghị Châu Â u về Thị giác Máy tính (ECCV), 740–755. Springer.

Liu, W .; Anguelov, D.; Erhan, D.; Szegedy, C.; Cây sậy, S .; Fu, C.-Y .; và Berg, AC 2016. Ssd: Máy dò multibox bắn một lần. Trong Hội nghị Châu Â u về Thị giác Máy tính, 21–37 . Springer.

Lưu, X.; Có thể.; Shi, Z .; và Chen, J. 2019. GridDehazeNet: Mạng đa tỷ lệ dựa trên sự chú ý để khử mùi hình ảnh. Trong Kỷ yếu của Hội nghị Quốc tế IEEE về Thị giác Máy tính (ICCV).

Loh, YP; và Chan, CS 2019. Làm quen với hình ảnh ánh sáng yếu với tập dữ liệu tối riêng. Thị giác Máy tính và Hiểu biết Hình ảnh, 178: 30–42.

McCartney, EJ 1976. Quang học của khí quyển: tán xạ bởi các phân tử và hạt. Newyork.

Mosleh, A.; Sharma, A.; Onzon, E.; Mannan, F.; Robidoux, N.; và Heide, F. 2020. Tối ưu hóa end-to-end phần cứng-in-the-end của các đường ống xử lý hình ảnh camera. Trong bài phát biểu của Hội nghị IEEE / CVF về Nhận dạng Mẫu và Thị giác Máy tính (CVPR), 7529–7538.

Narasimhan, SG; và Nayar, SK 2002. Tầm nhìn và bầu không khí. Tạp chí Quốc tế về Thị giác Máy tính, 48 (3): 233–254.

Polesel, A.; Ramponi, G.; và Mathews, VJ 2000. Hình ảnh xuất hiện thông qua mặt nạ không bị thay đổi thích ứng. Giao dịch IEEE về Xử lý hình ảnh, 9 (3): 505–510.

Tần, X.; Wang, Z .; Bai, Y .; Xie, X.; và Jia, H. 2020. FFA Net: Mạng lưới chú ý tổng hợp tính năng cho hình ảnh đơn lẻ de hazing. Trong Kỷ yếu của Hội nghị AAAI về Trí tuệ Nhân tạo , tập 34, 11908–11915.

Redmon, J .; Divvala, S.; Girshick, R .; và Farhadi, A. 2016. Bạn chỉ nhìn một lần: Phát hiện đối tượng hợp nhất, thời gian thực. Trong Kỷ yếu của IEEE / CVF Conference Computer Vision Pat tern Recognition (CVPR), 779–788.

Redmon, J .; và Farhadi, A. 2017. YOLO9000: tốt hơn, nhanh hơn, mạnh hơn. Trong Kỷ yếu của IEEE / CVF Conference Computer Vision Pattern Recognition (CVPR), 7263–7271.

Redmon, J .; và Farhadi, A. 2018. Yolov3: Một cải tiến gia tăng. arXiv: 1804.02767.

Ren, S.; Anh ấy, K .; Girshick, R .; và Sun, J. 2015. Nhanh hơn r-cnn: Hướng tới phát hiện đối tượng trong thời gian thực với các công trình mạng đề xuất vùng. Những tiến bộ trong hệ thống xử lý thông tin thần kinh, 28: 91–99.

Sindagi, VA; Oza, P.; Yasarla, R .; và Patel, VM 2020. Phát hiện đối tượng thích ứng miền dựa trên trước cho các điều kiện sương mù và mưa. Trong Hội nghị Châu Â u về Thị giác Máy tính (ECCV), 763–780. Springer.

Vương, G.; Guo, J .; Chen, Y. Li, Y. và Xu, Q. 2019. Một chiến lược học tập dựa trên PSO và BFO được áp dụng cho R-CNN nhanh hơn để phát hiện đối tượng trong lái xe tự hành. IEEE Access, 7: 18840–18859.

Vương, W .; Chen, Z .; Nhân dân tệ, X.; and Guan, F. 2021. Một phương pháp tăng cường hình ảnh ánh sáng yếu thích ứng. Trong Hội nghị liên quốc gia lần thứ mười hai về các hệ thống xử lý tín hiệu, tập 11719, 1171902. Hiệp hội Quang học và Photon quốc tế .

Bạn, S .; Tân, RT; Kawakami, R .; Mukaigawa, Y. và Ikeuchi, K. 2015. Mô hình hạt mưa kết dính, loại bỏ ngòi nổ trong video. Giao dịch IEEE về Phân tích Mẫu và Trí tuệ Máy móc, 38 (9): 1721–1733.

Yu, R .; Liu, W .; Zhang, Y. Qu, Z; Zhao, D.; và Zhang, B. 2018. Depexposure: Học cách phơi sáng các bức ảnh với phương pháp học đối phương được củng cố không ngừng. Trong Kỷ yếu của Hội nghị Quốc tế lần thứ 32 về Hệ thống Xử lý Thông tin Thần kinh (NeurIPS), 2153–2163.

Yu, Z .; và Bajaj, C. 2004. Một phương pháp nhanh chóng và thích ứng để tăng cường độ tương phản tuổi im. Trong Hội nghị Quốc tế về Xử lý Hình ảnh năm 2004. ICIP'04., Tập 2, 1001–1004. IEEE.

Zeng, H.; Cai, J .; Li, L.; Cao, Z .; và Zhang, L. 2020. Tìm hiểu bảng tra cứu 3D thích ứng với hình ảnh để nâng cao hiệu suất ảnh cao trong thời gian thực. Giao dịch IEEE trên Phân tích Pat tern và Máy thông minh.

Zhang, S.; Tuo, H.; Hu, J .; và Jing, Z. 2021. YOLO thích ứng tên miền để phát hiện tên miền chéo một giai đoạn. arXiv: 2106.13939.

Zhao, Z.-Q; Zheng, P.; Xu, S.-t .; và Wu, X. 2019. Phát hiện đối tượng với học sâu: Đánh giá. Giao dịch IEEE trên Mạng Nơ-ron và Hệ thống Học tập, 30 (11): 3212–3232.

Vật liệu bổ sung

Thiết kế bộ lọc Defog

Được thúc đẩy bởi phương pháp trước kênh tối thông thường (He, Sun và Tang 2009), chúng tôi thiết kế một bộ lọc khử mờ với một tham số. Trong mô hình tán xạ khí quyển (McCartney Năm 1976; Narasimhan và Nayar 2002), sự hình thành của một bóng tối hình ảnh có thể được xây dựng như sau:

$$I(x) = J(x)\,t(x) + A\,(1 - t(x)) \quad \text{trong} \tag{12}$$

đó I (x) là hình ảnh sương mù và J (x) đại diện cho cảnh rạng rỡ (hình ảnh rõ nét). A là ánh sáng khí quyển toàn cầu, và t (x) là ánh xạ truyền trung bình.

Để khôi phục hình ảnh rõ nét J (x), điều quan trọng là phải thu được ánh sáng khí quyển toàn cầu A và bản đồ sứ mệnh trung gian t (x). Để đạt được điều này, trước tiên chúng tôi tính toán bóng tối bản đồ kênh và chọn 1000 pixel sáng nhất hàng đầu. Sau đó, A được ước tính bằng cách lấy trung bình 1000 pixel này trong hình ảnh sương mù Tôi (x). Từ Eq. (12), chúng ta có thể suy ra rằng

$$\frac{I^{\text{Tối}}_{C}(x)}{A_C} = t(x)\frac{J_C(x)}{A_C} + (1 - t(x)) \tag{13}$$

trong đó C tặng kênh màu RGB. Bằng cách mất hai phút hoạt động, một trên các kênh và một trên bản vá cục bộ, trong phương trình trên, chúng ta có thể thu được:

$$\min_{C}\left(\text{phút}_{y\in\Omega(x)}\frac{I^{\text{Tối}}_{C}(y)}{A_C}\right) = t(x)\,\text{phút}_{C}\left(\text{phút}_{y\in\Omega(x)}\frac{J_C(y)}{A_C}\right) + (1 - t(x)) \tag{14}$$

Dựa trên kênh tối trước đó, chúng tôi có thể nhận được điều đó

$$J^{\text{tối}}(x) = \text{phút}_{C}\left(\text{phút}_{y\in\Omega(x)} J(y)\right) = 0 \tag{15}$$

Vì AC luôn dương nên phương trình. (15) có thể được viết là:

$$\min_{C}\left(\text{phút}_{y\in\Omega(x)}\frac{J_C(y)}{A_C}\right) = 0 \tag{16}$$

Bằng cách thay thế Eq. (16) thành Eq. (13), chúng ta có thể lấy:

$$t(x) = 1 - \text{phút}_{C}\left(\text{phút}_{y\in\Omega(x)}\frac{I^{\text{Tối}}_{C}(y)}{A_C}\right) \tag{17}$$

Chúng tôi giới thiệu thêm một tham số ω để kiểm soát mức độ xóa mờ. Có:

$$t(x, \omega) = 1 - \omega\,\text{phút}_{C}\left(\text{phút}_{y\in\Omega(x)}\frac{I^{\text{Tối}}_{C}(y)}{A_C}\right) \tag{18}$$

Vì thao tác trên có thể phân biệt được, chúng tôi có thể tối ưu hóa ω thông qua lan truyền ngược để làm cho bộ lọc defog linh hoạt hơn để phát hiện hình ảnh có sương mù.

Thí nghiệm

Thử nghiệm về Hình ảnh Sương mù Chúng tôi so sánh phương pháp của chúng tôi với đường cơ sở YOLOv3 (Redmon và Farhadi 2018), Defog + Detect (Hang et al. 2020; Liu et al. 2019), điều chỉnh chính (Hnewa và Radha 2021), và đa nhiệm vụ học tập (Huang, Le và Jaw 2020). Đối với phương pháp tiếp cận thích ứng miền, chúng tôi sử dụng công cụ dò tìm thích ứng miền đa quy mô một giai đoạn DAYOLO (Hnewa và Radha 2021) với nhiều đường dẫn thích ứng miền và các đường dẫn phân loại chính tương ứng ở các quy mô khác nhau của YOLOv3. Chúng tôi đặt giảm trọng lượng λ = 0,1 để đào tạo và mỗi đợt có 2 hình ảnh, một từ miền nguồn và một từ mục tiêu miền. Các siêu tham số khác được đặt giống như trong bản gốc.

Hình 7 cho thấy một số ví dụ trực quan về IA-YOLO của chúng tôi phương pháp, đường cơ sở YOLOv3 II và Defog + Detect các phương pháp. Cả GridDehaze (Liu et al. 2019) và MS BDN (Hang et al. 2020) đều có thể giảm hiệu ứng khói mù, điều này nói chung là có lợi cho việc phát hiện. Phương pháp IA-YOLO của chúng tôi không chỉ giảm khói mù mà còn nâng cao hình ảnh địa phương gradient, dẫn đến hiệu suất phát hiện tốt hơn.

Hình 8 cho thấy hai ví dụ về cách thức mà mod ule CNN-PP dự đoán các thông số của DIP, bao gồm cả thông số chi tiết giá trị và hình ảnh được xử lý bởi mỗi bộ lọc phụ. CNN PP có thể tìm hiểu một tập hợp các tham số DIP cho mỗi hình ảnh theo độ sáng, màu sắc, tông màu và thời tiết cụ thể thông tin. Sau khi hình ảnh đầu vào được xử lý bởi Mô-đun DIP, nhiều chi tiết hình ảnh được tiết lộ, phù hợp với nhiệm vụ phát hiện tiếp theo.

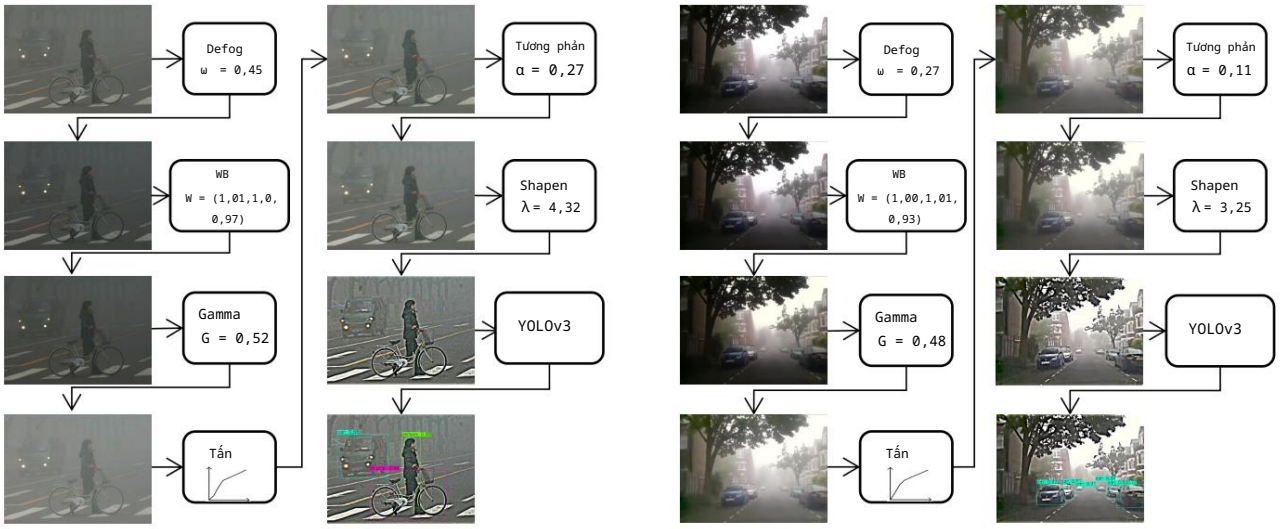
Thử nghiệm về hình ảnh trong điều kiện ánh sáng yếu Tổng số hình ảnh trong VOC_norm_trainval, VOC_norm_test và Ex Dark_test lần lượt là 12334, 3760 và 2563. Số lượng các thể hiện được liệt kê trong Bảng 6.

Chúng tôi so sánh phương pháp đã trình bày của chúng tôi với đường cơ sở YOLOv3, Nâng cao + Phát hiện (Guo và cộng sự 2020), DAYOLO, và DSNet trên ba bộ dữ liệu thử nghiệm. Hình 9 cho thấy các ví dụ trực quan trực quan về phương pháp IA-YOLO của chúng tôi, đường cơ sở YOLOv3 II và các phương pháp Nâng cao + Phát hiện . Nó có thể quan sát thấy rằng cả Zero-DCE (Guo et al. 2020) và IA YOLO đều có thể làm sáng hình ảnh và hiển thị chi tiết hình ảnh. IA-YOLO được đề xuất có thể làm tăng thêm độ tương phản của hình ảnh đầu vào, điều cần thiết để phát hiện đối tượng.

Phân tích hiệu quả

Trong khuôn khổ IA-YOLO được đề xuất của chúng tôi, chúng tôi giới thiệu một mô-đun tìm hiểu về CNN-PP vào YOLOv3, đây là một mô-đun nhỏ mạng chứa năm lớp phức hợp và hai lớp đầy đủ các lớp kết nối. Bảng 7 cho thấy phân tích hiệu quả của một số phương pháp được sử dụng trong các thí nghiệm của chúng tôi. Các phương pháp không được liệt kê được xác thực bằng kiến trúc YOLOv3. Cột thứ hai liệt kê số lượng tham số bổ sung qua mô hình YOLOv3. Cột thứ ba liệt kê thời gian chạy trên hình ảnh có độ phân giải 544 × 544 × 3 với một chiếc Tesla V100 duy nhất GPU. Có thể thấy rằng IA-YOLO chỉ thêm 165K có thể huấn luyện được thông số so với YOLOv3 trong khi vẫn đạt được hiệu suất tốt nhất trong tất cả các thử nghiệm với thời gian chạy tương đương. Lưu ý rằng IA-YOLO có ít thông số có thể huấn luyện hơn YOLO_deep II nhưng thời gian chạy của nó lâu hơn. Điều này là do quá trình nhập bộ lọc trong mô-đun DIP phải tính toán thêm.

Hình 7: Kết quả phát hiện bằng các phương pháp khác nhau trên ảnh sương mù RTTS thế giới thực. Từ trái sang phải: YOLOv3 II, GirdDehaze + YOLOv3 I, MSBDN + YOLOv3 I và IA-YOLO của chúng tôi. Phương pháp được đề xuất học cách giảm khói mù và nâng cao hình ảnh tương phản, dẫn đến hiệu suất phát hiện tốt hơn với ít phát hiện sai và bỏ sót hơn.



Hình 8: Các ví dụ về mô-đun DIP đã học và các đầu ra lọc của chúng. Mô-đun xử lý thích ứng hình ảnh có thể xuất ra các thông số bộ lọc tương ứng theo độ sáng, màu sắc, tông màu và thông tin thời tiết của từng hình ảnh đầu vào, để có được hiệu suất phát hiện tốt hơn.

Dataset	người	xe đạp	ô tô	xe buýt	xe máy	thuyền	chai	mèo	ghế	chó	Tổng cộng
Voc_norm_trainval 13256	1064	3267	822	337	1052	1140	1764	1593	3152	2025	29135
Voc_norm_test 4528	1201	213	418	919	164	325	263	469	358	756	489	8939
ExDark_test 2235						242	515	433	425	609	490	6450

Bảng 6: Thống kê các bộ dữ liệu đã sử dụng.

Hình 9: Kết quả phát hiện của các phương pháp khác nhau trên hình ảnh VOC_Dark_test tổng hợp (hàng trên cùng), ExDark_test trong thế giới thực trong điều kiện ánh sáng yếu hình ảnh (hai hàng dưới cùng). Từ trái sang phải: YOLOv3 II, ZeroDCE + YOLOv3 I và IA-YOLO của chúng tôi. Phương pháp đề xuất học cách làm cho hình ảnh sáng hơn với nhiều chi tiết hơn, dẫn đến hiệu suất phát hiện tốt hơn với ít sai sót và sai sót hơn sự phát hiện.

Phương pháp	Tốc độ tham số bổ sung (mili giây)	
YOLOv3	/	31
YOLOv3_deep II	412 nghìn	35
ZeroDCE	79 nghìn	34
MSBDN	31 triệu	94
GridDehaze	958 nghìn	51
IA-YOLO (Của chúng tôi)	165 nghìn	44

Bảng 7: Phân tích hiệu quả của các phương pháp so sánh.