



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)
دانشکده مهندسی برق
پروپوزال پروژه نهایی
بینایی ماشین (Machine Vision)

ارائه رویکردی به منظور تشخیص و ردیابی اشیاء با
استفاده از یادگیری عمیق

نگارش

محمد مهدی نجفی زاده

سارا سالمی

علی دانشپور

استاد درس

دکتر زهرا سادات شریعتمداری مرتضوی

آذر ۱۴۰۳

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

چکیده

تشخیص و ردیابی اشیاء یک امر حیاتی در بینایی ماشین با کاربردهای متنوع در حوزه‌های مختلف از جمله نظارت بر ترافیک، اتوماسیون صنعتی، خودروهای خودران و ... است. پیشرفت‌های اخیر در یادگیری عمیق به‌ویژه شبکه‌های عصبی کانولوشن (CNN)، دقت تشخیص و ردیابی اشیاء را به طور قابل توجهی بهبود بخشیده است. در این پژوهش قصد داریم یک سیستم ردیابی شیء با استفاده از الگوریتم‌های پیشرفته مانند YOLO، Faster R-CNN و SSD ارائه نمائیم. در این پژوهش ابتدا سناریوهای تشخیص و ردیابی تک شیء را بررسی کرده و سپس با گسترش قلمرو پژوهش، به بررسی پیچیدگی‌های سیستم‌های ردیابی چندشیء پرداخته و تلاش می‌کنیم رویکردی مناسب برای آن ارائه دهیم. به منظور ارزیابی مدل ارائه شده از مجموعه داده‌های معیار استاندارد مانند KITTI، MOT، و UAVDT استفاده می‌کنیم. همچنین از معیارهای ارزیابی مانند دقت و IoU برای ارزیابی دقت و توانایی رویکرد ارائه شده استفاده می‌کنیم. هدف اصلی این پژوهش کمک به توسعه سیستم‌های قابل اعتماد به منظور تشخیص و ردیابی اشیاء برای برنامه‌های کاربردی در دنیای واقعی است.

واژه‌های کلیدی:

بینایی ماشین، ردیابی اشیاء، پردازش تصویر، تشخیص اشیاء، یادگیری عمیق

فهرست مطالب

صفحه

عنوان

۱	۱	مقدمه
۱	۲	پیشینه پژوهش
۳	۳	معرفی برخی از روش‌های روش‌های موجود
۴	۴	مجموعه داده‌های موجود
۵	۵	مراحل عملیاتی پروژه
۷		منابع و مراجع

شکل	فهرست اشکال	صفحه
۱	معماری شبکه YOLOv1	۳
۲	سیر زمانی نسخه‌های مختلف الگوریتم YOLO	۴
۳	یک نمونه از خروجی الگوریتم YOLO	۴

صفحه	جدول	فهرست جداول
۵	۱	مجموعه داده‌های موجود

مقدمه

دنبال کردن اشیاء در ویدئو یکی از چالش‌های اصلی در بینایی ماشین است که در بسیاری از کاربردهای عملی از جمله نظارت امنیتی، تجزیه و تحلیل رفتار و ردیابی در محیط‌های صنعتی استفاده می‌شود. این فرآیند شامل شناسایی، مکان‌یابی و پیگیری مداوم اشیاء در طول یک توالی ویدئویی است. با پیشرفت‌های اخیر در یادگیری عمیق و به‌ویژه الگوریتم‌های شناسایی شیء مانند YOLO و Faster R-CNN، دقت و سرعت سیستم‌های شناسایی و ردیابی به طرز چشم‌گیری بهبود یافته است. این پیشرفت‌ها امکان پیگیری دقیق‌تر و سریع‌تر اشیاء را فراهم کرده و به کاربردهای پیچیده‌تری نظیر ردیابی چندشیء در محیط‌های شلوغ و متغیر گسترش یافته است.

کاربردهای عملی دنبال کردن اشیاء بسیار گسترده هستند. به‌عنوان مثال، در سیستم‌های امنیتی، این تکنولوژی برای ردیابی افراد یا خودروهای خاص به کار می‌رود. در صنعت، می‌تواند برای نظارت بر خطوط تولید و شناسایی نواقص به کار گرفته شود. در پزشکی، ردیابی اشیاء می‌تواند برای پیگیری حرکت ابزارهای پزشکی یا بیماران در بیمارستان‌ها مفید باشد. از دیگر کاربردها می‌توان به نظارت ترافیکی، تحلیل رفتار جمعیت، و ردیابی اشیاء در بازی‌های ورزشی اشاره کرد. از آنجا که پروژه ابتدا با پیگیری یک شیء آغاز می‌شود، این امکان فراهم می‌شود که الگوریتم‌ها ابتدا برای سناریوهای ساده بهینه شوند و سپس به سناریوهای پیچیده‌تر با چند شیء گسترش یابند. این گسترش در نهایت می‌تواند به سیستم‌های پیچیده‌تری منجر شود که قادر به پیگیری و تحلیل رفتار چندین شیء همزمان در محیط‌های دینامیک و پر پیچ‌وخمی مانند ترافیک شهری یا محیط‌های صنعتی هستند.

پیشینه پژوهش

سیستم‌های ردیابی اشیاء در ویدئو از زمان‌های دور تاکنون دستخوش تحولات زیادی شده‌اند. در ابتدا، روش‌های کلاسیک با استفاده از الگوریتم‌های ساده‌تری برای شناسایی و پیگیری اشیاء در ویدئوها استفاده می‌شدند. با پیشرفت علم یادگیری ماشین، روش‌های مبتنی بر ویژگی‌های دستی جای خود را به الگوریتم‌های پیچیده‌تری دادند که از قابلیت‌های یادگیری و تعمیم بیشتر برخوردار بودند. در ادامه، با ظهور یادگیری عمیق، روش‌های مدرن‌تری برای شناسایی و ردیابی اشیاء با دقت و سرعت بالا توسعه یافتند.

۱ - روش‌های کلاسیک (قبل از ۲۰۱۰):

در اوایل پژوهش‌ها در زمینه ردیابی اشیاء، الگوریتم‌های کلاسیکی مانند فیلتر کالمن و الگوریتم Hungarian برای پیگیری اشیاء استفاده می‌شدند. این روش‌ها عمدتاً بر اساس مدل‌های آماری و فیلتر کردن داده‌ها برای پیش‌بینی موقعیت اشیاء در فریم‌های بعدی عمل می‌کردند. به دلیل محدودیت‌های این الگوریتم‌ها در مواجهه با مشکلاتی مانند انسداد اشیاء یا تغییرات سریع محیط، کارایی آن‌ها در سناریوهای پیچیده

و محیط‌های دینامیک پایین بود.

۲ - روش‌های مبتنی بر ویژگی‌های دستی (۲۰۱۵-۲۰۱۰)

پس از آن، روش‌هایی مبتنی بر یادگیری ماشین معرفی شدند که از ویژگی‌های دستی تعریف‌شده برای شناسایی و پیگیری اشیاء استفاده می‌کردند. الگوریتم‌هایی مانند SVM و Random Forest در این دوره به کار گرفته شدند. این روش‌ها به دلیل وابستگی به ویژگی‌های از پیش تعریف‌شده و همچنین مشکلاتی در تعمیم‌پذیری، به تدریج جای خود را به مدل‌های پیچیده‌تر دادند.

۳ - ظهور شبکه‌های عصبی کانولوشنی (۲۰۱۷-۲۰۱۵)

از سال ۲۰۱۵ به بعد، با ظهور شبکه‌های عصبی کانولوشنی (CNN)، تحولی بزرگ در شناسایی و ردیابی اشیاء به وجود آمد. این شبکه‌ها می‌توانند ویژگی‌های پیچیده‌ای را از داده‌های ورودی استخراج کنند و دقت بالاتری در شناسایی و پیگیری اشیاء ارائه دهند. مدل‌هایی مانند YOLO و SSD [۳] در این دوره معرفی شدند که قابلیت شناسایی و ردیابی اشیاء در زمان واقعی را فراهم می‌کردند.

۴ - روش‌های مبتنی بر شبکه‌های پیشنهادی مناطق (۲۰۱۷-۲۰۱۹)

در این دوره، مدل‌هایی مانند Faster R-CNN [۲] معرفی شدند که از شبکه‌های پیشنهادی مناطق (RPN) برای شناسایی اشیاء استفاده می‌کنند. این مدل‌ها توانستند دقت بسیار بالاتری را در شناسایی اشیاء پیچیده و جزئیات ریز در ویدئوها ارائه دهند. با این حال، به دلیل پیچیدگی بالا و زمان پردازش طولانی‌تر، این مدل‌ها بیشتر برای سناریوهای پیچیده و دقیق مورد استفاده قرار می‌گرفتند.

۵ - روش‌های چندشیء و پیگیری هویت (۲۰۱۹ تاکنون)

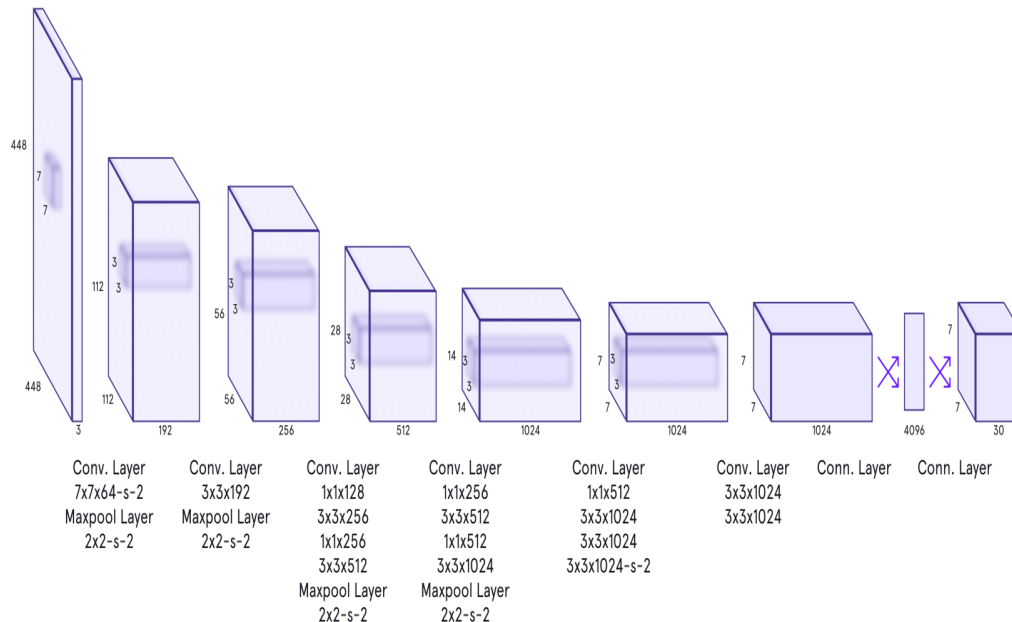
با پیشرفت‌های اخیر در زمینه پیگیری چندشیء، الگوریتم‌هایی مانند SORT (Simple Online and Realtime Tracking) و DeepSORT (Deep Learning-based SORT) [۵] معرفی شدند که به پیگیری هویت‌های مختلف در سناریوهای پیچیده کمک می‌کنند. این مدل‌ها علاوه بر شناسایی، هویت اشیاء را در طول ویدئوها تطبیق می‌دهند و قادر به پیگیری همزمان چندین شیء در محیط‌های شلوغ هستند. [۷]

معرفی برخی از روش‌های موجود

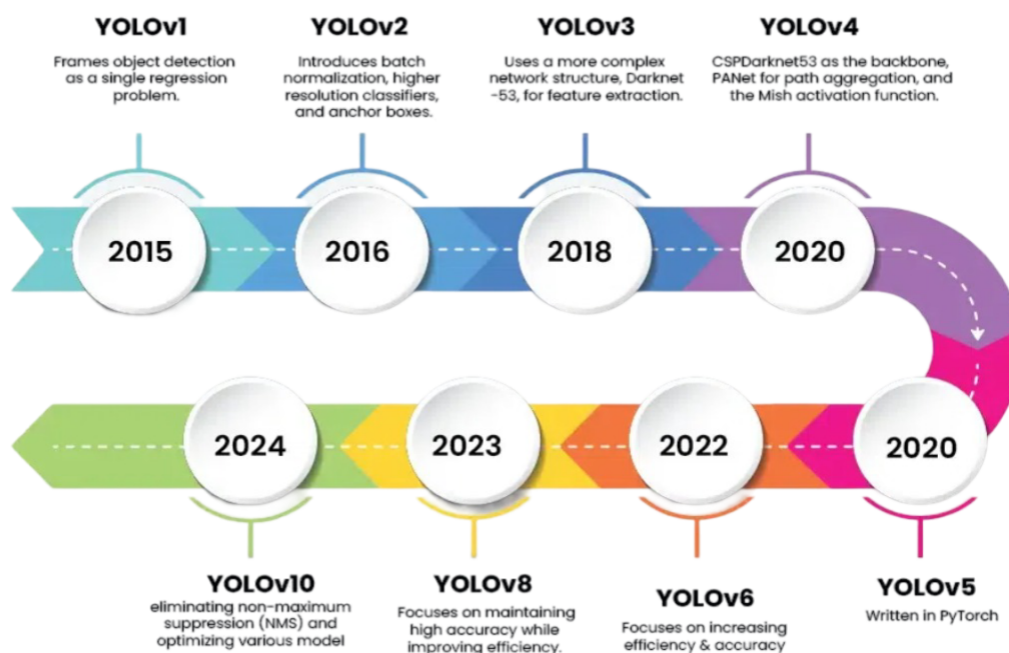
الگوریتم YOLO [6] (You Only Look Once):

برای شناسایی اشیاء در تصاویر از یک رویکرد مبتنی بر شبکه‌های عصبی کانولوشنی (CNN) استفاده می‌کند. ابتدا تصویر ورودی به یک شبکه تقسیم می‌شود که هر بخش از شبکه مسئول شناسایی یک یا چند شیء است. در هر سلول این شبکه، الگوریتم پیش‌بینی‌هایی شامل مختصات جعبه‌های محصور (Bounding Boxes) و احتمال دسته‌بندی اشیاء را انجام می‌دهد. به این صورت که هر سلول، یک جعبه محصور پیش‌بینی شده و احتمال مربوط به آن را می‌دهد.

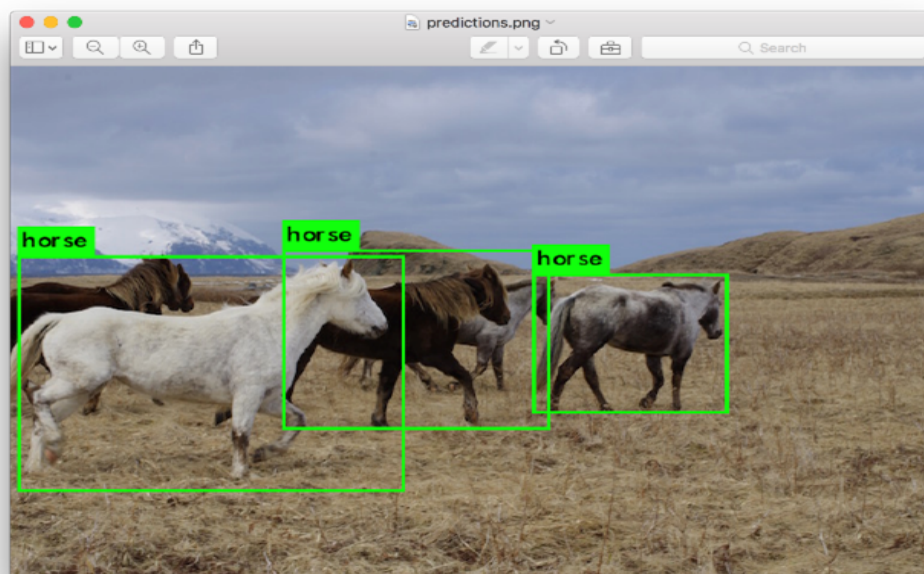
سپس از تکنیک‌هایی مانند Non-Maximum Suppression برای فیلترکردن جعبه‌های محصور تکراری و انتخاب بهترین پیش‌بینی استفاده می‌شود. عملکرد الگوریتم به این شکل است که تمام اشیاء را به طور همزمان در یک مرحله شناسایی می‌کند و نیاز به پردازش جداگانه برای هر شیء ندارد، که باعث افزایش سرعت آن می‌شود. تاکنون نسخه‌های مختلفی از این الگوریتم از YOLOv1 تا YOLOv10 تا ارائه شده است. در شکل ۲ سیر زمانی نسخه‌های مختلف معرفی شده و همچنین برخی از مهم‌ترین ویژگی‌ها مربوط به هر کدام مشخص است. همچنین در شکل ۳ یک نمونه از خروجی الگوریتم YOLO مشخص است.



شکل ۱: معماری شبکه YOLOv1



شکل ۲: سیر زمانی نسخه‌های مختلف الگوریتم YOLO



شکل ۳: یک نمونه از خروجی الگوریتم YOLO

مجموعه داده‌های موجود

در این قسمت برخی از مهم‌ترین مجموعه داده‌های موجود در این حوزه را معرفی می‌کنیم. به منظور ارزیابی رویکرد پیشنهادی این پروژه، از یک یا تعدادی از این مجموعه داده‌ها استفاده می‌کنیم.

جدول ۱: مجموعه داده‌های موجود

سال انتشار	کاربرد	حجم داده ها	پایگاه داده
۲۰۱۲	خودرویی، پیاده‌روها	متوسط	KITTI [۱]
۲۰۱۳	تشخیص علائم رانندگی	متوسط	GTSDB
۲۰۱۶	ردیابی چندشیء	متوسط	MOT16 [۴]
۲۰۱۷	ردیابی پیچیده	بالا	MOT17
۲۰۱۹	پهپادی	متوسط	UAVDT
۲۰۲۰	محیط‌های واقعی	بالا	TrackingNet
۲۰۲۱	نظارت شهری	بالا	AICity

مراحل عملیاتی پروژه

۱ - بررسی روش‌های مختلف تشخیص و ردیابی شیء

در ابتدا، باید روش‌های مختلف ردیابی شیء مورد بررسی قرار گیرند. این شامل تحلیل الگوریتم‌های معروف مانند YOLO [۶]، Faster R-CNN [۲]، و SSD [۳] است. همچنین، باید تفاوت‌ها و مزایا و معایب هر روش در شرایط مختلف محیطی و داده‌ای مشخص شود تا انتخاب روش مناسب برای پیاده‌سازی مشخص گردد.

۲ - پیاده‌سازی ساده از روش‌های مختلف

در مرحله بعد، پیاده‌سازی ساده از روش‌های منتخب باید صورت گیرد. این پیاده‌سازی باید به صورت آزمایشی و با استفاده از مجموعه داده‌های استاندارد انجام شود. هدف این است که عملکرد هر یک از روش‌ها در شرایط پایه بررسی و ارزیابی گردد.

۳ - بررسی چالش‌ها و توسعه به چند شیء

پس از پیاده‌سازی اولیه، باید بررسی‌های عمیقی در مورد چالش‌های موجود در ردیابی چندشیء و عواملی که بر کیفیت ردیابی تأثیر می‌گذارند، مانند همپوشانی اشیاء، تغییرات در پس‌زمینه، و تغییرات در محیط انجام شود. این مرحله شامل توسعه سیستم برای شناسایی و ردیابی چند شیء به طور همزمان و بررسی مشکلات خاص هر سناریو است.

۴ - بررسی معیارهای ارزیابی

در این مرحله، باید معیارهای ارزیابی مختلف مانند دقت، صحت، سرعت پردازش و توانایی الگوریتم در شرایط واقعی را بررسی کنیم. این معیارها باید برای ارزیابی عملکرد الگوریتم‌ها در موقعیت‌های مختلف تعیین و آزمایش شوند. معیارهایی مانند Intersection over Union (IoU) و Multiple Object Tracking Accuracy (MOTA) می‌توانند در این مرحله مفید باشند.

۵ - بررسی امکان پیاده‌سازی نرم‌افزاری و سخت‌افزاری

پس از ارزیابی الگوریتم‌ها، باید بررسی کنیم که آیا این الگوریتم‌ها قابلیت پیاده‌سازی در سیستم‌های نرم‌افزاری و سخت‌افزاری (مانند FPGA یا پردازنده‌های خاص) دارند یا خیر. این مرحله شامل ارزیابی نیازهای سخت‌افزاری و میزان تطابق الگوریتم‌ها با سیستم‌های بلادرنگ است.

۶ - ارتقاء روش پیشنهادی

در نهایت، باید به ارتقاء روش‌های پیشنهادی و بهینه‌سازی آن‌ها بپردازیم. این شامل بهبود الگوریتم‌ها برای افزایش دقت و سرعت، کاهش پیچیدگی‌ها، و انطباق با شرایط واقعی است. همچنین، در صورت نیاز، می‌توان از تکنیک‌های یادگیری تقویتی یا سایر الگوریتم‌های بهینه‌سازی برای بهبود عملکرد استفاده کرد.

منابع و مراجع

- [1] Geiger, Andreas, Lenz, Philip, Stiller, Christoph, and Urtasun, Raquel. Vision meets robotics: The kitti dataset. *International Journal of Robotics Research (IJRR)*, 2013.
- [2] Girshick, Ross. Fast r-cnn. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1440–1448, 2015.
- [3] Liu, Wei, Anguelov, Dragomir, Erhan, Dumitru, Szegedy, Christian, Reed, Scott, Fu, Cheng-Yang, and Berg, Alexander C. *SSD: Single Shot MultiBox Detector*, page 21–37. Springer International Publishing, 2016.
- [4] Milan, Anton, Leal-Taixe, Laura, Reid, Ian, Roth, Stefan, and Schindler, Konrad. Mot16: A benchmark for multi-object tracking, 2016.
- [5] Parico, Addie Ira Borja and Ahamed, Tofael. Real time pear fruit detection and counting using yolov4 models and deep sort. *Sensors*, 21(14), 2021.
- [6] Redmon, Joseph, Divvala, Santosh, Girshick, Ross, and Farhadi, Ali. You only look once: Unified, real-time object detection. pages 779–788, 06 2016.
- [7] Soleimanitaleb, Zahra, Keyvanrad, Mohammad Ali, and Jafari, Ali. Object tracking methods:a review. In *2019 9th International Conference on Computer and Knowledge Engineering (ICCKE)*, pages 282–288, 2019.