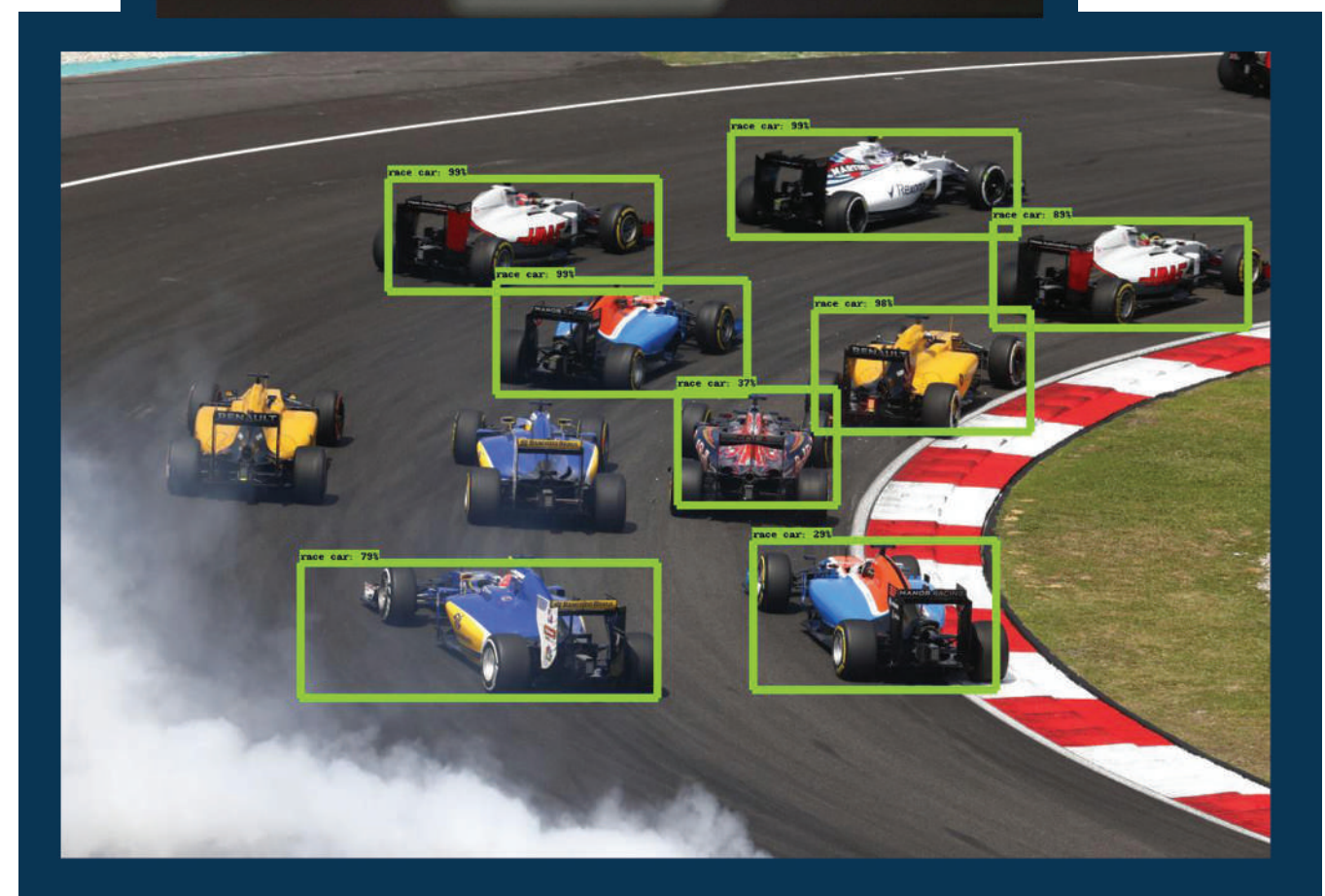
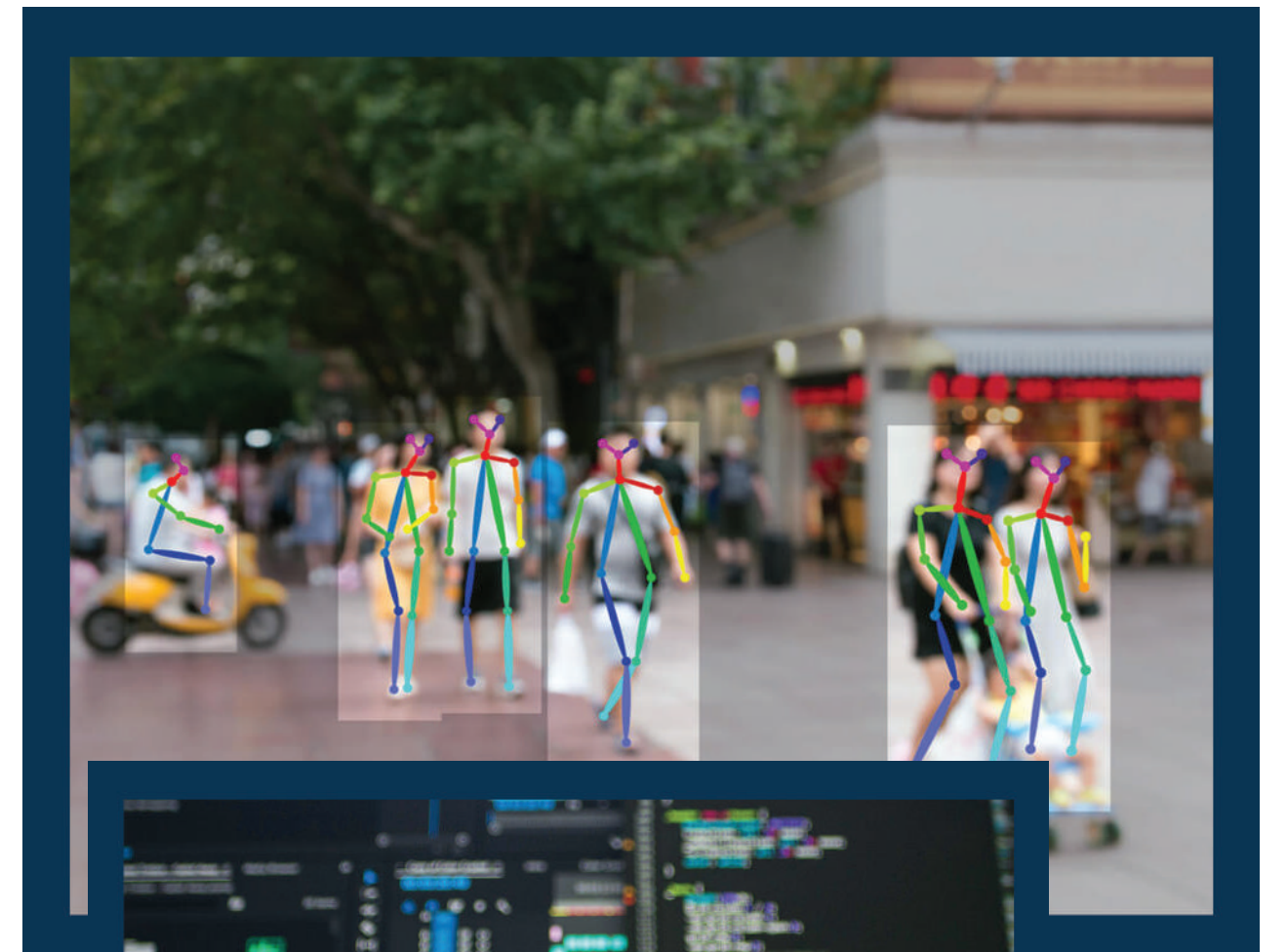


نشریه علمه

دنیای صفروویک



شماره 8 - بهمن 98





مقدمه ای بر بینایی ماشین

بازی و کنترلرها: یکی از نمونه های بارز آن دستگاه کینکت

مایکروسافت است.

نضارت: دوربین های مدار بسته عمومی برای تشخیص رفتارهای مشکوک.

خودروهای هوشمند: بینایی برای ماشین هم منبع اصلی اطلاعات است. خودروها قادر خواهند بود نور، علامت های ترافیکی و بقیه اشیا را تشخیص بدهند.

تکنیک های پرکاربرد

طبقه بندی تصاویر - Image Classification

در این تکنیک الگوریتمی می خواهیم که بتواند تصاویر را به دسته های متمایز تقسیم کند. برای حل این گونه مسائل محققان بینایی ماشین یک روش داده محور را پیشنهاد می کنند. به جای بررسی کدهای تک تک دسته بندی ها، آنها ابتدا برای کامپیوتر تعداد زیادی مثال تعریف می کنند، حال با توسعه الگوریتم های یادگیری، ماشین با بررسی مثال ها، ویژگی های هر دسته بندی را یاد می گیرد.

به عبارت دیگر ابتدا برای ماشین یک مخزن با مثال های تمرینی از تصاویر برچسبدار تعریف می کنیم و سپس ماشین هر برچسب را بررسی می کند.

ردگیری شیء - Object Tracking

ردگیری اشیا کاربردهای زیادی در ویدئوها و تصاویر متحرک دارد؛ جایی که مشاهده حرکات یک شی به صورت پیوسته

مقدمه

امروزه بینایی ماشین یکی از داغ ترین زمینه های تحقیقاتی در رابطه با یادگیری عمیق است. بینایی ماشین محل تلاقی علوم مختلفی مانند علوم کامپیوتر (گرافیک، الگوریتم ها، نظریه های ماشین، و معماری)، ریاضیات (بازیابی اطلاعات، یادگیری ماشین)، مهندسی ها (رباتیک، پردازش تصویر و ...)، فیزیک (نورشناسی)، زیست شناسی (علوم اعصاب) و روان شناسی (علوم شناختی) است.

"بینایی ماشین یعنی بیان توصیفاتی روشن

و معنی دار از اشیا فیزیکی در تصاویر"

(از کتاب "بالارد و براون"، 1982)

جاذبیت های این بینایی ماشین

تقریباً برای همه متخصصان حوزه IT روشن است که کاربردهای بینایی ماشین به سرعت در حال افزایش است. چندی از کاربردهایی که ممکن است روزانه با آن ها سر و کار داشته باشید:

تشخیص چهره: اسنپ-چت و فیسبوک از الگوریتم های تشخیص چهره برای اعمال فیلترها و تشخیص افراد در تصاویر استفاده می کنند.

بازیابی تصاویر: گوگل روزانه از الگوریتم های تحلیل تصاویر استفاده می کند.

سیگنال هایی که تشکیل دهنده تصویر یک شی هستند که توسط حسگرها دریافت شده است. البته این تعریف کلی نمیتواند همه جنبه های موضوع را نشان دهد پس به تفکیک در مورد هریک از زیر رویکردهای بینایی ماشین صحبت می کنیم.

الف-مطابقت دو سویی: در این رویکرد تصویر توسط دو دوربین دیده میشوند. این دو میتوانند از شبکه عصبی یکسان یا متفاوتی برای تحلیل تصویر استفاده کنند ضمناً فاصله این دو با یکدیگر و با تصویر نیز باید تنظیم شود. تصویری نهایی حاصل مقایسه تصویرهای این دو دوربین است (مشابه آنچه در جنگل تصادفی در یادگیری ماشین اتفاق می افتد به نوعی تجمیع نظرات برای بالا بردن دقت اعمال میشود).

ب-بازسازی صحنه: فرآیند ساخت و یا تکمیل یک تصویر توسط شبکه عصبی شاید مشکل ترین زیر مجموعه بینایی ماشین باشد. در این زمینه سعی میشود قسمتی از تصویر که به هر دلیلی از دست رفته است بازسازی شود. این رویکرد در زمینه تحلیل صحنه جرم و ساختن جهان های مجازی بسیار سودمند است.

ج-تشخیص اشیا: فرآیند شناسایی یک شی درون تصویر زمینه سوم از مبحث بینایی ماشین است. شرایط نوری متغیر در پیکسل ها میتواند معیار خوبی برای تشخیص شیء در یک تصویر دو بعدی باشد.

در نهایت شبیه به یک معادله خطی میشود) به هر ورودی سعی میکند ورودی ها را به خروجی های متناظر داده شده نگاشت کند. بار اول ضرایب بصورت رندوم یا تصادفی تنظیم میشوند اما از دفعات بعد با شیوه های ریاضیاتی (عموماً گرادیان کاهشی) سعی در تنظیم ضرایب بهتر میکند. تا جایی که میزان خطا (که خود میتواند معیارهای متفاوتی داشته باشد اما بصورت کلی میتوان فاصله مقدار خروجی واقعی و خروجی حاصل از تنظیم ضرایب را در نظر گرفت) از حد مشخصی کمتر شود و اینجاست که ضرایب مدنظر بعنوان ضرایب مطلوب میتوانند برای تولید خروجی های جدید بکار بسته شوند (برای بهینه سازی عملکرد پرسپترون و مفهوم و چگونگی اجتناب از over variance و over fit به منابع مربوطه مراجعه گردد) اما از عملکرد تکی هر پرسپترون مهمتر، عملکرد آنها در تعامل با یکدیگر است. خروجی هر پرسپترون میتواند ورودی دیگری باشد گاهی چند پرسپترون خروجی های خود را بعنوان ورودی به یکی میدهند تا آن ضرایب را روی آنها اعمال کند. بخصوص در مبحث بینایی ماشین نحوه چینش پرسپترون ها در هر لایه و نحوه ارتباط آنها برای تشخیص دقیق تر تصویر بسیار حائز اهمیت است.

بینایی ماشین اهداف و کاربردها

شرح آنچه که در مورد عملکرد بینایی در انسان داده شد را در مورد ماشین میتوان بدین صورت خلاصه کرد: پردازش



سید علی آله طه

بینایی ماشین و شبکه های عصبی کانولوشنی

را در شبکه چشم تشکیل میدهد که همان چیز است که شما میبینید. تمام موج های مذکور اصطلاحاً یک سیگنال هستند که مغز بعنوان جعبه سیاه مساله با پردازش صحیح آنها میتواند فرد را قادر به تشخیص تصویر کند. حال اینگونه در نظر بگیرید که تمام موجها اعم از مرئی یا نامرئی، صوتی یا تصویری فرکانس و سایر مشخصات مربوط به خودشان را دارند که با پردازش صحیح آنها میتوان آنچه در مغز اتفاق میفتد را برای ماشین شبیه سازی کرد.

شبکه های عصبی کانولوشنی

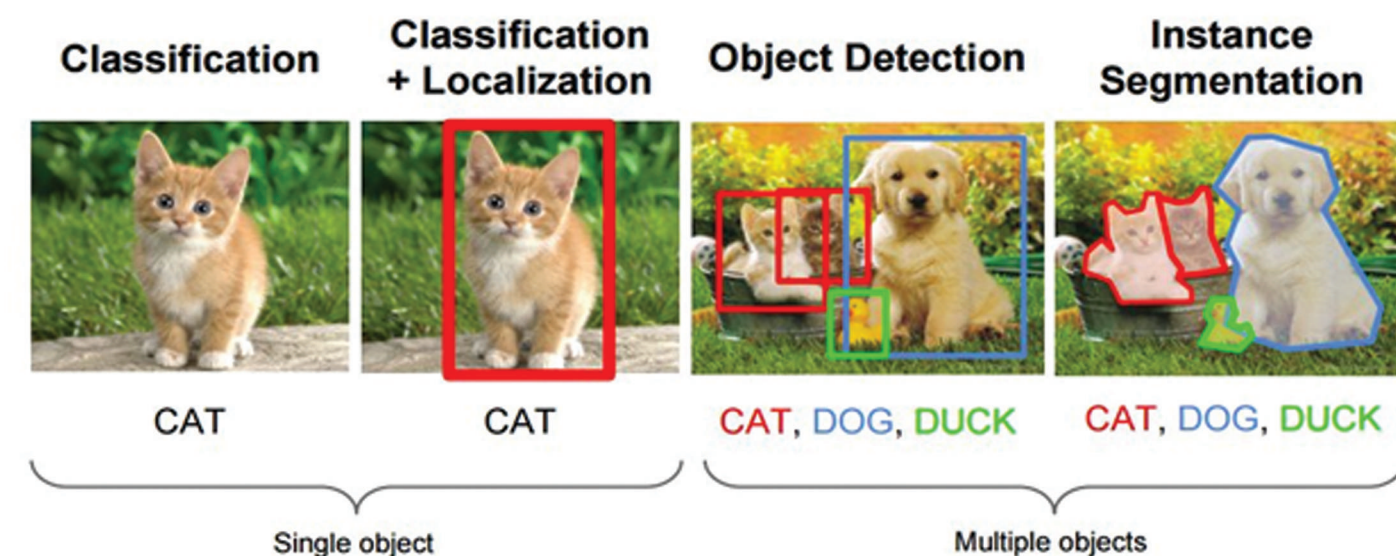
اکنون میخواهیم جعبه سیاه را باز کنیم تا به جزئیات بیشتری از عملکرد آن برسیم. به راستی شبکه های عصبی چه میکنند که تصویر را با دقت خوبی تشخیص میدهند؟ کوچکترین عضو هر شبکه عصبی پرسپترون (Perceptron) است. پرسپترون میتواند هر تعداد ورودی و خروجی داشته باشد اما آنچه که در نتیجه نهایی کار مهم است اولاً پردازش هر پرسپترون روی ورودی-هایش است ثانیاً نحوه تعامل پرسپترون ها با یکدیگر (در ساختار لایه ای شبکه های عصبی) است. برای آنکه به جزئیات بیشتری از نحوه کار پرسپترون پی ببریم ناچاریم کمی به یادگیری ماشین مراجعه کنیم. برای تمرین دادن یک پرسپترون تعدادی از ورودی ها و خروجی-های متناظر لازم است. پرسپترون با تنظیم کردن ضرایب و اضافه کردن یک عدد تحت عنوان بایاس (که

احتمالاً کسانیکه درس سیگنال ها و سیستم ها را داشته اند با واژه "کانولوشن" آشنا باشند اما اگر جزو دسته ای هستید که این درس را پاس نکرده اید نگران نباشید. Convolution در لغت به معنای پیچش است اما اجازه دهید قبل از پرداختن به تعریف علمی کانولوشن و شبکه های عصبی کانولوشنی به بیان اهمیت آن در مباحث بینایی ماشین بپردازیم.

نحوه عملکرد مغز

مفهوم کانولوشن یک مفهوم بسیار اساسی در پردازش سیگنال ها اعم از صوتی و تصویری است. برای آنکه هدف عملکرد کانولوشن روشن تر شود ابتدا آن را بصورت یک جعبه سیاه (Black Box) در نظر میگیریم و ورودی ها و خروجی های آن را بررسی میکنیم لذا نحوه عملکرد مغز را بهترین مثال در این جهت میدانم. در نظر بگیرید هنگامی که شما به یک تصویر نگاه میکنید آن تصویر تصویری است که در شبکه چشم شما تشکیل شده است اما همه میدانیم که اگر مغز نباشد انسان قادر به دیدن نیست؛ مغز است که پردازش تصویری که پیش روی-تان است را بر عهده دارد پس میتوان مغز را به عنوان جعبه سیاه مذکور در نظر گرفت. هنگامی که نور به تصویر میرسد همه پیکسل های آن تصویر نور را با طول موجی که بصورت منحصر به فرد مربوط به رنگ آن پیکسل است بازتاب می کنند و این نورهای بازتاب شده وقتی به مغز می رسند مغز مجدداً تمام آنها را کنار هم میگذارد و تصویری

Computer Vision Tasks



مورد نیاز است. امروزه یکی از تکنیک های لازم برای ماشین های خودران ردگیری است و شرکت های بزرگی مانند تسلا و اوبر روی این تکنیک کار می کنند. یکی از روش ها برای پیاده سازی ردگیری روش متمایزکننده است.

در این روش روی هر فریم از ویدئو، شی مورد نظر را از پس زمینه جدا میکنیم سپس با استفاده از الگوریتم های یادگیری عمیق تصاویر جداسازی شده را در همه فریم ها تشخیص میدهیم. سپس از الگوریتم هایی مانند شبکه های عصبی کانولوشنی (CNN) برای تشکیل رد اشیا استفاده می کنیم.

دانشگاه صادر کننده مجوز:

دانشگاه تهران

صاحب امتیاز:

انجمن علمی مهندسی کامپیوتر

پردیس فاربی دانشگاه تهران

زمینه انتشار:

علمی - تخصصی کامپیوتر

مدیر مسئول:

علی داودآبادی فراهانی

سر دبیر:

محمد خوشنود

طراح و صفحه آرا و گرافیک:

سید محمد مهدی هاشمی

احمد قمی

شماره چاپ:

شماره هشتم

1398

هرکدام از رویکردها ضعف و قوت هایی دارند اما آنچه مهم است با بکار بردن هر کدام سر جای خود و یا ترکیب آنها با یکدیگر میتوان به نتایج عالی دست یافت. برای مثال تشخیص پلاک یک نمونه نسبتا ساده از کاربردهای بینایی ماشین میباشد. شبکه عصبی باید قادر باشد در یک تصویر (احيانا شلوغ) یک مستطیل شامل چند عدد به نام پلاک را تشخیص دهد (مورد ج) سپس هر عدد نوشته شده روی آنرا به تفکیک تشخیص دهد (مورد ج) سپس با توجه به تمرینی که از قبل داده شده است باید تشخیص دهد که آن چه عددیست و یا اگر مخدوشی و زدگی در تعدادی از اعداد وجود دارد باید بتواند به کمک رویکرد بازسازی صحنه آن را با دقت بالا ترمیم کند.

سخن آخر

بدون استفاده از شبکه های عصبی کانولوشنی بینایی ماشین معنایی نخواهد داشت زیرا بینایی ماشین تماما بر پردازش سیگنال مبتنی است. در یک دهه گذشته بینایی ماشین تبدیل به یک حوزه بالغ در هوش مصنوعی و بطور دقیق تر یادگیری عمیق شده است. علت اقبال محققان و سرمایه گذاران شرکت های بزرگ صنعتی و تجاری به این زمینه چیزی جز اتوماتیک شدن کارهایی که با نظارت انسان صورت می گرفته است نیست.



The poster for MVIP 2020 features a blue and white color scheme. At the top, it displays the CIVILICA logo on the left and the University of Tehran logo on the right. The title "MVIP 2020" is prominently displayed in the center. Below the title, the text "انجمن بینایی ماشین و پردازش تصویر ایران" (Iranian and the 1st International Conference on Machine Vision and Image Processing) is written in Persian. The English title "The 11th Iranian and the 1st International Conference on Machine Vision and Image Processing" is also present. The poster includes a central graphic of an eye with circuitry, a robot head, and a code symbol. It lists the submission deadline as 21 Dec. 2019, acceptance notification as 15 Jan. 2019, and camera ready due as 25 Jan. 2019. The location is specified as "پرديس فارابی دانشگاه تهران، قم، ایران" (University of Tehran, College of Farabi, Qom, Iran). The dates "۳۰ بهمن و ۱ اسفند ۱۳۹۸" (19 & 20 February 2020) are also mentioned. Contact information for the faculty of engineering and the conference website is provided at the bottom.