چکیده:

امروزه استفاده از سامانه های هوشمند رواج گسترده ای در دنیا پیدا کرده است. به خصوص بعد از معرفی هوش مصنوعی و پردازش تصویر که جزء جدا ناپذیر هوش مصنوعی می باشد تمامی افراد و شرکت‌ها در تلاش هستند تا سامانه‌های خود را مجهز به این دو ابزار قدرتمند بکنند. یکی از موضوعات در حال رشد امروزه دنبال کردن(tracking) اشیا و یا حتی انسان ها برای اهداف خاصی می باشد. برای این کار می توان از ارسال و دریافت سیگنال های مختلف استفاده کرد ولی همواره این روش با مشکلات خاص خود روبه‌رو بوده است، از جمله این مشکلات تداخل سیگنال می باشد. برای حل این مشکلات امروزه به کمک پردازش تصویر و هوش مصنوعی وظیفه دنبال کردن اجسام انجام می گیرد. از جمله کاربرد های این کار می توان به مدیریت انبار داری در کارخانه ها، مدیریت منابع انسانی در محیط کار و یا حتی کاربرد های ورزشی از جمله دنبال کردن ورزشکار ها در حین مسابقه اشاره کرد. در این پروژه به کمک هوش مصنوعی (بخش نرم افزار) و یک ساختار دینامیکی (بخش سخت افزار) یه دنبال کننده (tracker) ساخته شده است که می تواند یک انسان را دنبال کند.

فصل اول: مقدمه

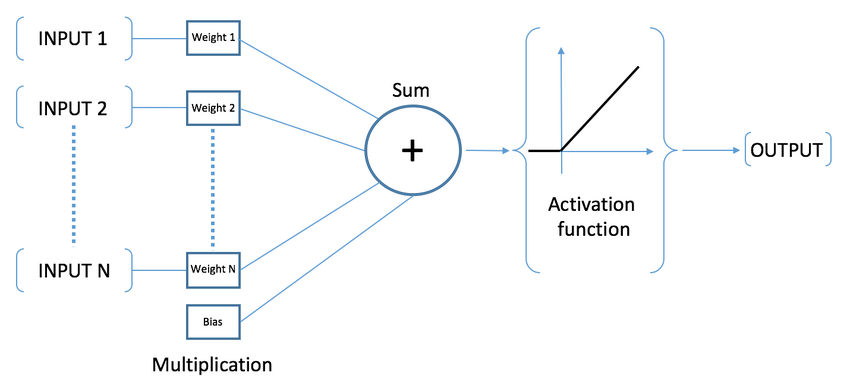
فصل دوم: پردازش تصویر وشبکه‌های عصبی

**1-1 مقدمه:** پردازش تصویر دیجیتال شامل دست‌کاری تصاویر با استفاده از رایانه‌های دیجیتال است. در دهه‌های گذشته استفاده از آن به‌طور تصاعدی افزایش‌یافته است. کاربردهای پردازش تصویر دیجیتال از پزشکی گرفته تا سرگرمی و سنجش‌ازدور است. رشتۀ پردازش‌تصویر دیجیتال یک رشتۀ گسترده است که شامل تکنیک‌های پردازش سیگنال دیجیتال و همچنین تکنیک‌های خاص تصاویر است.یک تصویر را می‌توان به‌عنوان یک تابع در نظر گرفت f(x,y) از دو متغیر پیوسته x و y . برای پردازش دیجیتالی که باید از آن نمونه‌برداری کرد و به ماتریسی از اعداد تبدیل شود. ازآنجاکه رایانه، اعداد را با استفاده از دقت محدود نشان می‌دهد، برای نمایش دیجیتالی، این اعداد باید کمی شوند. پردازش‌تصویر دیجیتال شامل دست‌کاری اعداد دقیق محدود است. پردازش تصاویر دیجیتالی را می‌توان به چنددسته تقسیم کرد: بهبود تصویر، ترمیم تصویر، تجزیه‌وتحلیل تصویر و فشرده سازی تصویر. پردازش تصویر دارای روش های مختلفی می باشد که آن هار را می توانیم در دو گروه اصلی تقسیم بندی بکنیم. گروه اول روش هایی می باشند که به کمک یک سری روابط ریاضی پیاده سازی شده اند. در این روش ها الگوریتم های ریاضی را بر روی تصویر پیاده سازی کرده و خروجی مورد نظر را دریافت می کنند. به عنوان مثال به کمک این روش ها می توان فضای رنگی تصویر را عوض کرد و یا اختلال1 تصویر را حذف کرد و یا حتی اختلال1 به آن اضافه کرد. یکی از معروف ترین کتاب‌خانه ها در این زمینه که به زبان پایتون نوشته شده است، کتاب‌خانه open-cv می باشد. گروه دوم روش های مبتنی بر یادگیری عمیق می باشند که از آن ها با عنوان شبکه های عصبی نیز یاد می شود. تفاوت روش های این گروه در ساختار شبکه عصبی می باشد که استفاده شده است. امروزه شبکه های عصبی فراوانی برای پردازش تصویر معرفی شده است که هر کدام دارای ویژگی خاص می باشد. به عنوان مثال شبکه های Retina Net و YOLO مثال هایی از شبکه های عصبی برای پردازش تصویر می باشند که برای تشخیص اشیا2 از آن ها استفاده می شود. به علت این که در این پروژه از شبکه های عصبی برای قسمت پردازش تصویر استفاده شده است در ادامه تاریخچه ای از پردازش تصویر و شبکه های عصبی بیان می شود.

**1-2 تاریخچه:** اواخر سال 1980 بود که شبکه­ های عصبی در علم پزشکی ظهور کرده و به کار گرفته شدند. شبکه­ های عصبی که دارای دو نوع یادگیری با معلم3 و بدون معلم4 می­باشند، هر دو به عنوان راه حل ­هایی موفقیت آمیز در پزشکی مورد استفاده قرار گرفتند.تشخیص بیماری ای که امکان نتیجه­ گیری از داده­های فراوان به دست آمده از آن به صورت دستی وجود ندارد، از مهم­ترین کمک ­های شبکه عصبی به علم پزشکی می­باشد. شبکه عصبی از جمله تکنولوژی­ هایی است که به شکل­های مختلف به پشتیبانی از خدمات پزشکی می پردازد. شناخت تصویر در شبکه عصبی علاوه بر کمک به پیشرفت درمان، موجب کاهش هزینه­ های ناشی از درمان، ارتباط موثر تر میان اطلاعات به دست آمده از شرایط بیمار، ارتقاء سطح کیفی خدمات بهداشتی و… می­گردد.حوزه پردازش تصویر و شناخت تصویر در شبکه عصبی در علم پزشکی بازه گسترده ­ای از کاربرد های گوناگون نظیر تشخیص دیابت چشمی با استفاده از تصاویر گرفته شده از شبکیه چشم، بخش بندی تصویرهای ام آر آی تهیه شده به منظور تشخیص وجود یا عدم وجود تومور در مغز انسان و… را در بر می­گیرد.اگر چه در گذشته برخی از روش­های تشخیص بیماری بر اساس استخراج ویژگی­ ها و اطلاعات از روی تصاویر پزشکی و به صورت کاملا دستی و با صرف وقت و انرژی فراوان به وسیله متخصصان صورت می پذیرفت، اما امروزه روشی جدید با هدف تشخیص و طبقه بندی تصاویر پزشکی بدون آن که نیازمند استخراج ویژگی ­ها به صورت دستی باشد و با به کارگیری شناخت تصویر در شبکه عصبی پایه گذاری شده است. مهم­ترین مزیت شناخت تصویر در شبکه­ عصبی، توانایی این شبکه­ ها در استخراج خودکار ویژگی­ های تصویر با به کارگیری مفهوم یادگیری عمیق می­باشد. همین مزیت مهم موجب استفاده از شناخت تصویر در شبکه عصبی در سال­ های اخیر شده است.به عنوان مثال در تشخیص بیماری رتینوپاتی دیابتی که از آن تحت عنوان دیابت چشمی نیز یاد می­شود، رگ­ های شبکیه موجود در چشم بیمار در اثر بیماری دیابت به منظور خون رسانی بهتر نسبت به حالت معمولی خود گشادتر شده و تغییر شکل می­دهند. شناخت تصویر در شبکه عصبی به منظور استخراج ویژگی­ های شبکیه چشم مورد استفاده قرار گرفته و مراحل پیشرفت بیماری با استفاده از این تصاویر معین می­گردد. تشخیص درست پولیپ­ ها در طی تصویر برداری کلونوسکپی، تشخیص خودکار رگ گرفتگی­ های ریوی، تصاویر سی تی اسکن، تشخیص سرطان سینه، تشخیص غده­های لنفاوی با استفاده از تصاویر سی تی اسکن، تشخیص غده­های لنفاوی شکمی، تشخیص خودکار آناتومی بدن و… از دیگر مزایای استفاده از شناخت تصویر در شبکه عصبی می­باشد.شناخت تصویر در شبکه عصبی علاوه بر موارد فوق الذکر، در زمینه­ اندازه ­گیری و بخش ­بندی تصاویر موجود در زمینه پزشکی نظیر بخش بندی لوزالمعده در تصاویر به دست آمده از اسکن توموگرافی، بخش بندی تصاویر گرفته شده از مغز کودکان، بخش بندی غشاء عصبی در تصاویر الکترومیکروسکوپی گرفته شده، بخش بندی تصاویر گرفته شده از غضروف­ های زانو در اسکن­ های ام آر آی، اندازه ­گیری ضخامت لایه عروق و رگ­ ها و…. نیز بسیار کاربرد دارد.

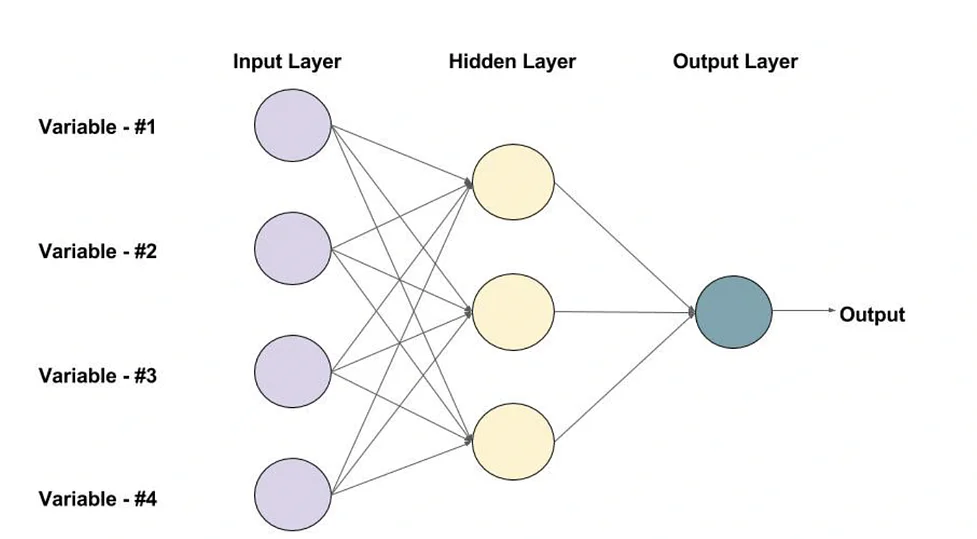
همان گونه که پیش­تر نیز اشاره کردیم، شبکه عصبی و شناخت تصویر در شبکه عصبی ایده­ای جذاب و در عین حال قدیمی است. مفهوم اولیه شبکه عصبی را نخستین بار Frank Rosenblatt در مقاله­ای که در سال 1957 به چاپ رساند، تحت عنوان پرسپترون معرفی نمود. وی در سال 1958 موفق به طراحی سیستم اولیه آنالیز یک تصویر ساده و شناسایی اشکال هندسی شد. اصلی ­ترین هدف Rosenblatt ساخت سیستمی پر کاربرد به منظور بخش بندی یا کلاس بندی تصاویر بود. این در حالی است که او علاقه فراوانی به درک بهتر کارکرد مغز از این راه بود. بعدها این علاقه Rosenblatt طرفداران زیادی را به خود جذب نمود! هر نورون موجود در یک شبکه عصبی تابعی ریاضیاتی است که مجموعه­ای از ورودی­ های دارای وزن را پردازش می­کند. لازم به ذکر است که در صورت بالا بودن وزن سیناپس، آن سیناپس خروجی نورون را به مقدار قابل توجه ­تری تحت تاثیر قرار می­دهد. خروجی نورون به تابعی غیر خطی تحت عنوان تابع فعال ارسال می­شود. این تابع توانایی مدل سازی پدیده­ های غیر خطی پیچیده را به شبکه می­دهد.یک شبکه با استفاده از تنظیم نمودن وزن سیناپس­ ها و بازخورد دریافت شده از عملکرد سیستم نسبت به ورودی­ های گوناگون موفق به یادگیری مدل می­گردد. کلاس بندی صحیح تصاویر توسط شبکه موجب افزایش مقادیر وزن­ های موثر در جواب و کاهش دیگر وزن­ها می شود. در صورتی که کلاس بندی تصویر به اشتباه صورت بپذیرد، کلیه وزن ­ها برعکس خواهند شد. نخستین شبکه ­های عصبی بدین شکل موفق به یادگیری شدند. در ادامه ساختار کلی شبکه های عصبی را بررسی کرده و روند آموزش و استفاده از این شبکه ها را بیان می کنیم.

**1-3 ساختار**: همان طور که در قسمت تاریخچه نیز به آن اشاره شد ساختار شبکه های شبیه به ساختار مغز انسان می باشند و همانند آن از یک سری سلول5 تشکیل شده اند که هر کدام از سلول ها در واقع بیانگر یک رابطه ریاضی می باشد. در شکل 1000 ساده ترین حالت یک سلول شبکه عصبی را مشاهده می کنیم.



شکل(1000):تصویر نمونه از نورون های شبکه عصبی

همانطور که در شکل 1000 مشاهده می کنیم هر سلول عصبی از سه قسمت اصلی تشکیل شده است: وزن های ورودی یا همان پارامتر های یادگیری (این وزن ها یا پارامتر ها حین آموزش شبکه تغییر می کنند)، عملگر جمع کننده و تابع فعال ساز. عملکرد هر سلول بدین گونه است که ورودی شبکه در وزن های سلول ضرب شده و وارد سلول می شود، سپس در سلول مقادیر با هم جمع شده و در نهایت از یک تابع فعال ساز عبور داده می شود. هدف از تابع فعال ساز ایجاد یک رابطه غیر خطی می باشد، همان طور که میدانیم داده های ورودی می توانند دارای ویژگی های خطی و یا غیر خطی باشند. حال برای تشکیل یک شبکه عصبی باید تعداد زیادی از تک سلول ها را با یک ساختار مناسب کنار هم قرار دهیم. در شکل 1001 ساده ترین حالت کنار هم قرار گرفتن سلول ها را مشاهده می کنیم.



شکل(1001): ساده ترین حالت شبکه های عصبی

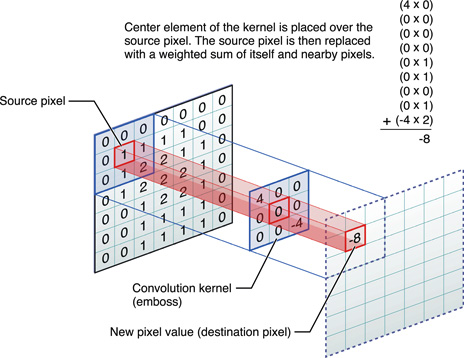
با توجه به شکل 1001 هر شبکه عصبی از یک لایه ورودی تشکیل شده است که بیان گر داده های ورودی است. با توجه به ابعاد داده های ورودی تعداد سلول های لایه ورودی متغیر می باشد. دقت کنیم که سلول های لایه ورودی دارای پارامتر یادگیری نمی باشند و فقط مبین داده های ورودی می باشند. بعد از لایه ورودی لایه های پنهان شبکه را مشاهده می کنیم که وظیفه بسیار مهم استخراج ویژگی ها را بر عهده دارند. سلول های این لابه دارای پارامتر های یادگیری(وزن های ورودی) می باشند که این پارامتر ها حین آموزش شبکه تغییر می کنند. تعداد لایه های پنهان با توجه به ساختار و کاربرد شبکه می تواند متغیر باشد. در نهایت بعد از لایه های پنهان لایه خروجی شبکه را داریم که با توجه به نوع خروجی مورد نظر تعداد سلول های آن می تواند متفاوت باشد. در شکل 1002 ساختار یک شبکه عصبی با تعداد لایه های پنهان بیشتر از یک را مشاهده می کنیم.



شکل(1002): ساختار کامل شبکه عصبی

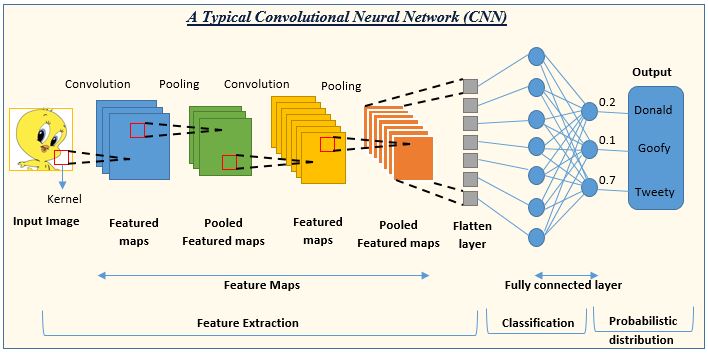
با توجه به شکل 1002 و توضیحات قبلی در مورد سلول ها و لایه های شبکه می توان عملکرد یک شبکه عصبی را بدین گونه توصیف کرد: در ابتدای هر شبکه یک لایه مربوط به داده های ورودی داریم که متناسب با بعد6 داده های ورودی، تعداد سلول های آن متفاوت می باشد. یک نکته بسیار مهم در مورد سلول های لایه ورودی این است که این سلول ها فقط بیانگر مقدار داده ورودی می باشند و نماینده هیچ گونه رابطه ریاضی نیستند و همچنین فاقد پارامتر های یادگیری می باشند. در ادامه لایه های پنهان قرار دارند که میتوان گفت مهمترین بخش شبکه های عصبی را تشکیل می دهند.در هر کدام از این لایه ها متناسب با ساختار شبکه تعداد سلول های متفاوتی قرار می گیرند. وظیفه این لایه ها استخراج ویژگی های داده های ورودی می باشد. می توان گفت اساس کار شبکه های عصبی، همین لایه های پنهان و ویژگی هایی است که در این لایه ها استخراج می شوند و بر اساس آن ها خروجی مورد نظر در لایه آخر ساخته می شود. بعد از لایه های پنهان لایه خروجی قرار دارد که همانند لایه های پنهان دارای پارامتر های یادگیری می باشند ولی تعداد سلول های موجود در آن وابسته به بعد داده های خروجی متغیر می باشد.تا این مرحله ساختار کلی شبکه های عصبی را مورد بررسی قرار دادیم، ولی با توجه به این که داده های مورد استفاده ما در این پروژه از نوع تصویر می باشند در ادامه ساختار شبکه های عصبی که متناسب با وظایف7 پردازش تصویر می باشد را بررسی می کنیم.

همان طور که می دانیم هر تصویر از سه بعد طول، عرض و کانال تشکیل شده است. هر تصویر دارای یک طول و عرض مشخصی می باشد و تعداد کانال های آن عموما سه می باشد که اصطلاحا به آن 7RGB می گویند. برای پردازش داده های تصویری در شبکه های عصبی به جای لایه های تک بعدی از سلول ها(همانند شکل1002) از لایه های با ساختار دو بعدی که سلول ها در دو بعد قرار گرفته اند استفاده می شود.ماهیت این سلول ها دقیقا مشابه ماهیت سلول های ساختار تک بعدی می باشد.اصطلاحا به این لایه های سلولی دو بعدی، کرنل8 می گویند.این کرنل ها بر روی تک تک پیکسل های تصویر ورودی اعمال می شوند و یک تصویر جدید را تولید می کنند در شکل 1003 یک نمونه از این کرنل ها را مشاهده می کنیم.



شکل(1003):نمونه ای از کرنل های مورد استفاده دز پردازش تصویر

با استفاده از این کرنل ها می توان تصویر با ابعاد و کانال های متفاوتی تولید کرد. البته دقت کنیم معمولا تغییر ابعاد تصویر اهمیت چندانی ندارد و در شبکه های عصبی تعداد کانال های تصویر را به کمک تعداد کرنل هایی که تعریف می کنند تغییر می دهند و بدین گونه ویژگی های داده های ورودی را استخراج می کنند.مهم ترین نکته در مورد کرنل ها این است که مقادیر درایه های این کرنل ها به عنوان پارامتر یادگیری در شبکه های عصبی تعریف می شوند و حین آموزش شبکه این پارامتر ها تغییر می کنند تا به بهترین مقدار خود دست پیدا کنند و بدین گونه است که با بهینه شدن مقادیر این کرنل ها میتوانیم ویژگی های مهم داده های ورودی را استخراج کرده و از آن ها استفاده کنیم. به این گونه لایه ها که از کرنل ها ساخته شده اند لایه های کانوولوشنی9 می گویند و همچنین به شبکه های عصبی که از این لایه ها ساخته شده اند، شبکه های عصبی کانوولوشنی10 گفته می شود. در شکل 1004 یک نمونه کامل از این شبکه های عصبی کانوولوشنی را مشاهده می کنیم.



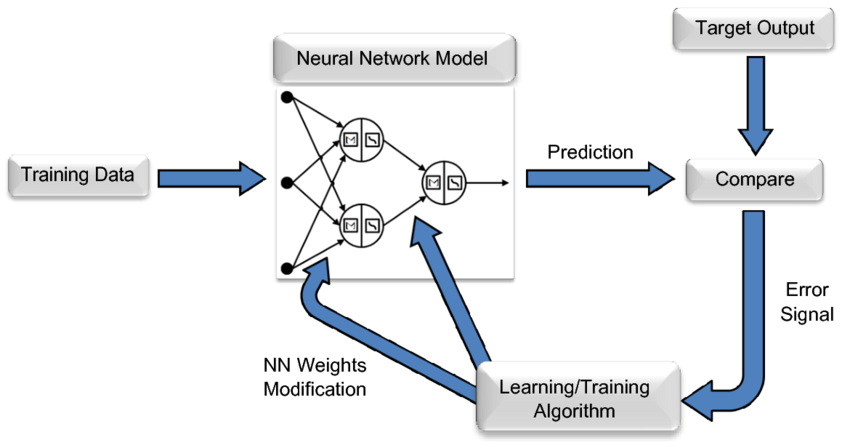
شکل(1004): نمونه کامل یک شبکه عصبی کانوولوشنی

همان طور که در شکل 1004 مشاهده می‌کنیم به کمک لایه های کانوولوشنی، لایه های پنهانی را تشکیل می دهیم تا ویژگی های داده ورودی که تصویر می باشد استخراج شوند. لایه ورودی همانند لایه ورودی شبکه های غیر کانوولوشنی همان داده ورودی(تصویر) می باشد و فقط متناسب با بعد تصویر تعداد سلول های آن می تواند متغیر باشد. در مورد لایه خروجی نکته ای که در شبکه های کانوولوشنی وجود دارد این است که با توجه به خروجی مطلوب و مورد نظر لایه خروجی می تواند شامل کرنل برای خروجی تصویر و یا شامل لایه های تک بعدی شبکه های معمولی برای خروجی عدد و یا غیر تصویر باشد. در ادامه با نحوه آموزش شبکه های عصبی آشنا می شویم و توضیح داده می شود که چگونه پارامتر های یادگیری تنظیم می شوند.

1-4 آموزش شبکه: در مورد آموزش شبکه های عصبی دو قسمت اصلی وجود دارد که شامل ساختار شبکه و داده می باشد.در مورد ساختار شبکه در قسمت های قبل توضیح داده شد.داده هایی که از آن ها برای آموزش استفاده می شود شامل دو قسمت می باشند: داده های ورودی و داده های خروجی. ابتدا داده های ورودی وارد شبکه می شوند و شبکه یک خروجی تولید می کند، خروجی تولید شده با خروجی واقعی مقایسه شده و بر اساس یک الگوریتم ریاضی به نام پس انتشار خطا10 و خطای تولیدی شبکه پارامتر های شبکه تنظیم می شوند.اساس کار الگوریتم پس انتشار خطا این می باشد که از خطای شبکه نسبت به هر کدام از پارامتر های شبکه که به آن ها اصطلاحا وزن های شبکه می گویند مشتق می گیریم و گرادیان آن را حساب می کنیم، سپس با یک ضریب کوچک(بین 0 و 1) این مقدار را از مقدار قبلی پارامتر کم می کنیم. بدین گونه پارامتر های شبکه به گونه ای تنظیم می شوند که به کمترین مقدار خطا در شبکه دست پیدا کنیم. رابطه 1000 بیانگر رابطه ریاضی الگوریتم پس انتظار خطا می باشد.

فرمول(1000) 

در شکل 1005 ساختار بلوک دیاگرامی آموزش شبکه را مشاهده می کنیم. با توجه به این شکل ماهیت الگوریتم پس انتشار خطا را در روند آموزش شبکه به خوبی مشاهده می کنیم.



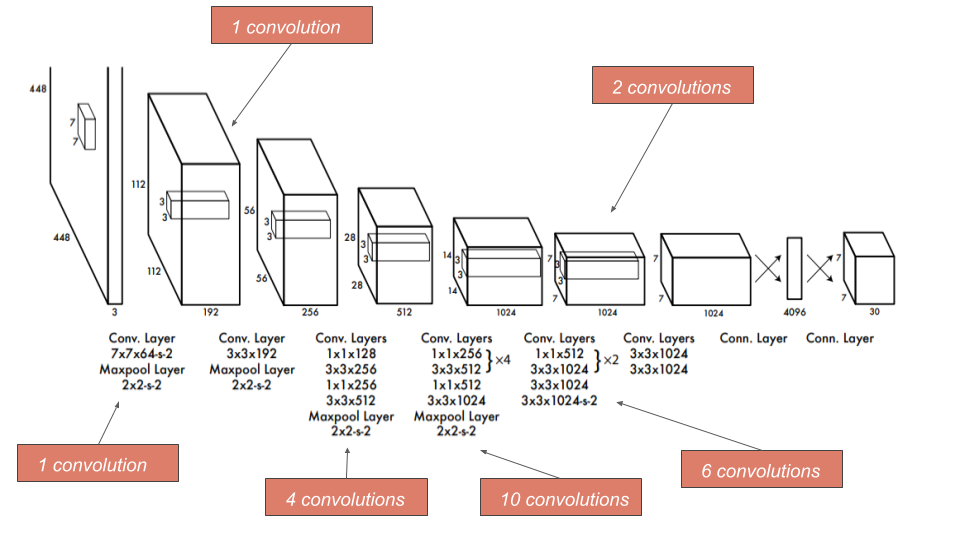
شکل(1005): ساختار بلوک دیاگرامی آموزش شبکه

تا این قسمت ساختار شبکه های عصبی را بررسی کرده و روند آموزش آن را توضیح دادیم. در ادامه دو مورد از ابزار های پردازش تصویری که در روند پروژه از آن استفاده کردیم را توضیح خواهیم داد.

1-5 ابزار های پردازش تصویر: با توجه به موضوع و هدف پروژه ما باید از ابزار پردازش تصویری استفاده کنیم که قادر به تشخیص یک شئ خاص باشد که اصطلاحا به آن Object Detection گفته می شود.در همین راستا ابزار های معرفی شده در این قسمت با هدف تشخیص اشیا پیاده سازی شده اند.در روند این پروژه از 2 ابزار پردازش تصویری شامل شبکه عصبی YOLO و الگوریتم HAAR Cascade در کتابخانه OpenCV استفاده شد که در نهایت به علت نتیجه بسیارخوبی که شبکه عصبی YOLO داشت در نهایت از این شبکه عصبی برای پروژه استفاده شد.

1-5-1 الگوریتم HAAR Cascade در کتابخانه OpenCV: تشخیص اجسام با استفاده از الگوریتم HAAR Cascade، یک روش بسیار موثر است که اولین بار توسط Paul Viola و Michael Jones در سال ۲۰۰۱ مطرح شد. این الگوریتم بر پایه یادگیری ماشین بوده و بدین صورت عمل می کند که با استفاده از تصاویری که چهره در آنها وجود دارد (تصاویر مثبت) و تصاویری که چهره در آنها وجود ندارد (تصاویر منفی) تابعی را آموزش می دهد تا بتواند مولفه هایی مانند وجود فاصله بین چشم ها را پیدا کند. این الگوریتم منحصر به تشخیص چهره نیست؛ و بیشتر یک آموزش دهنده است و می توان با آموزش دادن تابع دلخواه خودتان هر شی مانند ماشین، میز، مداد و … را شناسایی کنید. OpenCV مجموعه ای از توابع از پیش آموزش دیده را برای تشخیص چهره، چشمان، لبخند و … در خود دارد که می توان به سادگی از آنها استفاده کرد. مشکل اصلی این الگوریتم این بود که برای تشخیص شئ مورد نظر حتما باید شئ بدون هیچ مانعی(هر چند کوچک) در مقابل دوربین قرار می گرفت و اگر یک مانع بسیار کوچک در مقابل آن باشد دیگر قادر به تشخیص نیست و این مورد باعث کم شدن دقت انجام کار میشد. با استفاده از شبکه عصبی YOLO توانستیم مشکل موجود در الگوریتم HAAR را حل کنیم. در ادامه به توضیح شبکه عصبی YOLO می پردازیم.

1-5-2 شبکه عصبی YOLO : شبکه عصبی YOLO یک شبکه عصبی عمیق با لایه های کانوولوشنی می باشد. همان طور که در قسمت ساختار شبکه های عصبی آشنا شدیم از لایه های کانوولوشنی برای پردازش داده های تصویری استفاده می شود. همچنینی YOLO یک شبکه عمیق11 می باشد.اصطلاحا به شبکه های عصبی که از لایه های زیادی تشکیل شده و دارای پارامتر های بسیار زیادی می باشند شبکه های عمیق می‌گویند. در شکل 1006 ساختار شبکه YOLO را مشاهده می‌کنیم.



شکل(1006): ساختار شبکه عصبی YOLO