

گزارش مقاله

محمد حسین عبدی-۹۶۲۲۲۰۶۱

بینایی و تشخیص اجسام توسط انسان یکی از قابلیت‌های پیچیده و در عین حال شگفت‌انگیزی است که توسط مغز ما صورت می‌گیرد. این عمل نیاز به استخراج یک سری ویژگی از جسم مشاهده شده دارد که توسط ناحیه بینایی مغز صورت می‌گیرد. تصمیم‌گیری در باره اینکه در کدام دسته این جسم را جا بدهد نیز در لایه‌های بالاتر مغز انجام می‌شود. برای این امر دسته‌بندی یک سری از شبکه‌های عصبی اسپایک (Spike Neural Network) دخیل هستند و این امر را صورت می‌دهند.

برای شبیه‌سازی کردن کاری که این شبکه‌های عصبی صورت می‌دهند، یعنی استخراج ویژگی‌های بصری، می‌توانیم از STDP برای استخراج یک سری ویژگی‌های ابتدایی و پیش پا افتاده استفاده کنیم.

در این مطالعات از Shallow SNN ها استفاده شده و یک لایه یادگیری هم استفاده می‌شود.

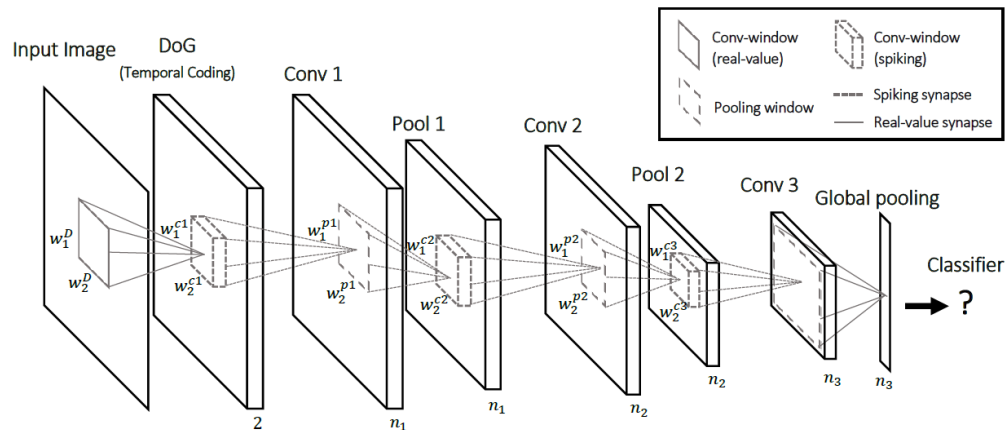
این تحقیقات از یک deep SNN استفاده کرده و در آن یک مدل LIP اسپایکی با ساختار پیچیده (convolusitonal) وجود دارد. به جای یادگیری‌های معمول که در یک لایه اتفاق می‌افتند در این شبکه‌ها یادگیری را در کل شبکه خواهیم داشت. از یک temporal coding استفاده کرده‌ایم و به این صورت عمل می‌کند که نورون‌های فعال‌تر زودتر عمل می‌کنند و زودتر فایر را انجام می‌دهند. به این دلیل که از STDP استفاده می‌کنیم نورون‌های مجموعه ما رفته رفته ویژگی‌های الگوی نمونه را یاد می‌گیرند و به همین دلیل نیاز به تعداد زیاد نمونه‌ها نیست.

شبکه یادگیری که در مغز وجود دارد، مانند شبکه‌های اسپایکی هستند و تفاوت این شبکه یادگیری با شبکه‌های معمولی در این است که نورون‌های این شبکه اسپایکی هستند و بیشتر از یک لایه قابل یادگیری دارند به همین دلیل تصاویر را در مقیاس بزرگ می‌توانند پردازش کنند.

نحوه جلوگیری از overfitting در این شبکه به این صورت است که چون داریم از STDP استفاده می‌کنیم پس مثل شبکه‌های معمول یادگیری supervised backpropagation نداریم و unsupervised عمل می‌کنیم.

این مقاله یک SDNN بر مبنای STDP پیشنهاد می‌کند که spiketime coding دارد. این شبکه همانطور که در عکس مشخص است از چند لایه convolutional و pooling تشکیل شده و در انتها نیز یک لایه برای کلاس‌بندی ورودی‌ها با توجه به میزان و نحوه فعالیت نورون‌ها در آخرین لایه pooling، کلاس‌ها را با استفاده از global receptive field مشخص می‌کند. در اینجا سعی شده لایه‌های convolutional عملکردی شبیه به simple cell ها در مغز داشته باشند و به دنبال استخراج

ویژگی‌های داده‌های ورودی هستند. لایه‌های pooling نیز مانند سلول‌های complex عمل می‌کنند و از بین داده‌های نورون-های دیگر عمل ماکسیمم‌گیری انجام می‌دهند.



ساختار شبکه به این صورت است که در ابتدا و بعد از گرفتن ورودی از یک فیلتر DoG عبور می‌کند که این فیلتر در ابتدا contrast را تشخیص می‌دهد و بر اساس آن و با استفاده از temporal coding یک کدگذاری مبنی بر میزان قدرت contrast به صورت اسپایک انجام می‌دهد. پس از آن اولین لایه convolutional را داریم که مانند ناحیه V1 در مغز عمل می‌کند. یعنی خطوط و زاویه خطوط را درک می‌کند. پس از آن یک لایه pooling که همه ویژگی‌های استخراج شده را در یک ناحیه جمع می‌کند و باعث می‌شود که حجم داده‌های تصویری کمتر بشود. لایه دوم convolutional و pooling بعدی مانند قسمت V2 و V4 در مغز عمل کرده که شکل‌ها و اشیاء را به طور کلی تشخیص می‌دهد. لایه آخر convolutional استفاده شده در شبکه، که در کنار یک لایه دیگر از pooling قرار دارد شبیه‌سازی قسمت IT مغز را انجام می‌دهد یعنی تشخیص می‌دهد که شی داده شده به عنوان ورودی چه شی‌ای است. برای مثال صورت یک انسان یا یک دوچرخه یا یک ماشین ... در نهایت هم یک global pooling صورت می‌گیرد تا داده‌ها را بتوان کلاس‌بندی کرد.

در لایه‌های convolutional هر نورون بر اساس وزنی که توسط سیناپس ورودی‌شان به آن‌ها می‌رسد انتخاب می‌شود. این انتخاب برای این است که هر نورون یک ویژگی بینایی مخصوص را از داده‌ها استخراج کند. این استخراج ویژگی خاص نیز به این صورت تعیین می‌شود که هر نورون در یک mapping خاص وجود داشته باشد ویژگی‌های مربوط به آن کلاس از نورون‌های خاص را استخراج می‌کند.

نوع نورون‌ها در لایه‌های convolutional به صورت nonleaky IF هست و اسپایک‌ها را از نورون قبلی می‌گیرند؛ وقتی پتانسیل داخلی آن‌ها به پتانسیل فایر شدن نورون رسید، اسپایک می‌زنند و به نورون بعدی می‌رود.

یک مکانیسم مهاری که در این شبکه وجود دارد به این صورت است که زمانی که نورون‌ها یک اسپایک ورودی می‌گیرند همانطور که در بالا اشاره شد هر نورون در یک مپینگ و نگاشت مشخص قرار دارد. مکانیسم مهاری به این صورت عمل می‌کند که وقتی

نورون در یک مکان خاص فایر را انجام بدهد نورون‌های دیگر مهار می‌شوند و از اسپایک زدن آن‌ها جلوگیری می‌شود تا زمانی که ورودی بعدی برسد و در نتیجه هر کدام از نورون‌ها تنها مجاز هستند یک مرتبه اسپایک بزنند.

فعالیت نورن‌ها در هر لایه با استفاده از فرمول زیر مشخص می‌شود:

$$V_i(t) = V_i(t-1) + \sum_j W_{ji} S_j(t-1)$$

که $V_i(t)$ در واقع نشان دهنده پتانسیل درونی نورون convolutional در زمان t است.

W_{ji} وزن سیناپسی است که بین نورون قبلی j و نورون بعدی i وجود دارد.

S_j اسپایک‌های نورون j ام را نشان می‌دهد. پس اگر عبارت $S_j(t-1)$ مساوی با یک باشد یعنی در لحظه $t-1$ نورون j ام اسپایک داشته.

نورون‌های pooling همه از نوع IF هستند و وزن ورودی سیناپس‌ها و آستانه آن‌ها همگی روی یک تنظیم شده است. پس اولین اسپایک ورودی آن‌ها را فعال کرده و منجر به اسپایک زدن می‌شود. در این لایه هیچ یادگیری اتفاق نمی‌افتد و هر نورون نیز حداکثر یک بار فایر می‌کند.

یادگیری فقط در لایه‌های convolutional اتفاق می‌افتد. این یادگیری به این صورت است که: در نورون‌ها یاد می‌گیرند که با ترکیب کردن اطلاعات ساده‌تری که از لایه‌های قبلی استخراج شده استفاده می‌کنند و اطلاعات بینایی پیچیده‌تری را می‌توانند استخراج کنند. این ترکیب کردن نسبت به وزنی که نورون‌ها دارند صورت می‌گیرد و وزن‌ها از قاعده STDP به دست می‌آید:

$$\Delta w_{ij} = \begin{cases} a^+ w_{ij} (1 - w_{ij}), & \text{if } t_j - t_i \leq 0, \\ a^- w_{ij} (1 - w_{ij}), & \text{if } t_j - t_i > 0 \end{cases}$$

این رابطه به این معنی است که در صورت فایر کردن نورون‌مان اگر همه ورودی‌های قبلی فایر کرده باشند، وزن نورون فایر شده افزایش می‌یابد.

نورون‌ها در یک صفحه مشترک باید وزن‌های مشابه هم داشته باشند. برای این کار در یک صفحه اولین نورونی که فایر می‌شود STDP را اعمال می‌کند و تغییرات وزنی که به وجود می‌آید را با تمام نورون‌های صفحه به اشتراک می‌گذارد.