

$$\text{Softmax}(\vec{x}_i) = \frac{e^{x_i}}{\sum_{j=1}^K e^{x_j}} \quad e^{x_i} = \begin{bmatrix} 4.08 & 7.61 & 8.42 & 2.72 & 6.8 \end{bmatrix}$$

0 1 2 3 4

$$\text{Softmax}(\vec{x}_i) = \begin{bmatrix} \frac{4.08}{29.63} & \frac{7.61}{29.63} & \frac{8.42}{29.63} & \frac{2.72}{29.63} & \frac{6.8}{29.63} \end{bmatrix} = \begin{bmatrix} 0.14 & 0.25 & 0.28 & 0.09 & 0.22 \end{bmatrix}$$

$$\text{Cross-Entropy Loss} = -\frac{1}{N} \sum_{j=1}^K p_i \log q_i \Rightarrow -\frac{1}{N} \sum_{j=1}^K y_n \log \hat{y}_n + (1 - y_n) \log (1 - \hat{y}_n) \quad (\text{ب})$$

$$\cancel{J} = \frac{1}{5} \sum_{i=1}^5 J_i \Rightarrow J = -\log(p_{\text{true}}) = -\log(0.09) = -(-2.41) = 2.41$$

2- (الف) لایه 1 ← ابعاد خروجی (3, 1024, 1024) و یا راستها 0 - لایه 2 ← ابعاد خروجی (16, 512, 512) و یا راستها 16 + (16 × 3 × 2) = 2868

لایه 3 ← ابعاد خروجی (16, 256, 256) و یا راستها 4 - لایه 4 ← (32, 252, 252) و یا راستها 12832

لایه 5 ← ابعاد خروجی (32, 126, 126) و یا راستها 0 - لایه 6 ← (64, 126, 126) و یا راستها 18496

لایه 8 ← (64, 63, 63) و یا راستها 0 - لایه 9 ← (256, 61, 61) و یا راستها 147712

لایه 12 ← (10, 10) و یا راستها 10250

2868 + 12832 + 18496 + 147712 + 203168 + 18250 + 491754

ب) لایه 2 ← 2 × 7 × 8 × 16 × 512 × 512 = 37,349,120 و یا راستها 4 - لایه 4 ← 16 × 3 × 3 × 32 × 252 × 252 = 486,400 و یا راستها 16,616,400

لایه 6 ← 32 × 3 × 3 × 64 × 126 × 126 = 294,053,376 و یا راستها 7 - لایه 7 ← 64 × 3 × 3 × 64 × 126 × 126 = 588,106,752

لایه 9 ← 64 × 3 × 3 × 256 × 61 × 61 = 172,755,456 و یا راستها 11 - لایه 11 ← 1024 × 256 = 262,144 و یا راستها 12 - لایه 12 ← 10 × 1024 = 10,240

فقط لایه ها 20 Gen و 20 Denge ضرب دارند. مجموع تقریبی 488,923,288.3

ج) لایه 20 Gen برای دردی عکس است و کرنل آن در بعدی است و سریع تر و سبک تر است.

در حالی که 30 Gen برای دیدن است و کرنل آن به بعدی است و می تواند درگیری های زمانی را هم استخراج کند.

3- داده افزایی بر روی مدل overfit می خورد چرا که با افزایش تعداد داده های آموزش مدل به الگوهای مندیبه در داده های افزایش یافته دست می یابد و عمیق آن بیشتر می شود. با افزایش لایه های شبکه از underfit جلوگیری می کنیم چرا که مدل می تواند مندیبه تر یاد بگیرد و الگوهای سخت تر را مندیبه شود. کاهش تعداد ویژگی های دردی یا dimensionality reduction هم برای مدل overfit استفاده می شود چون اگر داده کم باشد رنجیر زیاد به مشکل curse of dimension می خوریم یعنی تعداد داده ها نسبت به ویژگی ها کمتر است و مدل به سبب الگوهای مندیبه دردی مجموعه داده آموزش حفظ می شود.

مردار است همیشه مدل underfit است چرا که جنیت loss ها بالا است یعنی هیچ کدام از مجموعه داده ها را خوب یاد نگرفته. نمودار است درست مدل overfit است چرا که loss داده آموزش خیلی پایین است در حالی که validation بالا است یعنی مجموعه داده آموزش را حفظ کرده.

4- الف) در CNN اتصال محلی است یعنی هر $Node$ شبکه عصبی (نورون) به یک سری نورون همسایه متصل است در حالی که در MLP عبارتاً $fully connected$ است. در CNN یک اشتدادک وزنی داریم چرا که یک نرل در کل تصویر لغزش می‌کند.

در حالی که MLP نه ارد این مورد را. در CNN تعداد پارامتر کمتر است در حالی که در MLP این عدد بسیار است. در CNN اعتبار مکانی حفظ می‌شود در حالی که در MLP این اطلاعات از بین می‌رود. در نتیجه با توجه به توصیف بالا CNN ها برای تصاویر مناسب‌ترند.

ب) میدان تأثیر محدوده‌ای از تصویر است که یک نورون از لایه محلی تر به آن حساس است. میدان تأثیر هر چه بزرگ تر نورون توانا تر در یادگیری ویژگی‌های بزرگ مقیاس تر و هر چه کوچک‌تر یادگیری ویژگی‌های محلی.

پ) $max pooling$ مقدار بیشینه در ناحیه را برمی‌گزیند در حالی که $Avg pooling$ میانگین را. از آنجمله $max pooling$ مقدار به زیر بردن و برجسته کردن ویژگی‌ها است در حالی که $Avg pooling$ اطلاعاتی تر را حفظ می‌کند. البته $max pooling$ ممکن است اطلاعات ظریف را حذف کند. بلکه است ویژگی‌های مهم را نگه دارد که در طرف مقابل $Avg pooling$ کاربرد رایجی $max pooling$ در $object Detection$ است. کاربرد رایجی $avg pooling$ در توصیف کلی تعداد ویژگی‌های سبک تر است.

ت) افزایش $Kernel-Size$ باعث افزایش میدان تأثیر می‌شود و یادگیری ویژگی‌های بزرگ تر بیشتر می‌شود و همچنین می‌توانی علم بالایی رود.

با افزایش $stride$ ابعاد خروجی کاهش می‌یابد و حافظه کاهش می‌یابد سرعت بیشتر می‌شود اما جزئیات علم از بین می‌برد.

تعداد فیلترها اگر افزایش یابد ویژگی‌های استخراجی علم افزایش می‌یابد و قدرت مدل بیشتر می‌شود و حافظه بیشتری می‌شود و زمان پردازش کمتر زیاد می‌شود.