

الف) تعداد وزن ها در حالت FC ، $I \times W_2 \times H_2$ باشد در این صورت صر

نوردن لایر قبل به آنها لایر وصل است \rightarrow درکل $I \times W_2 \times H_2 \times J \times W \times H$

وزن داریم

تعداد وزن ها در حالت CNN ، هر کدام از I دیگری مربوط به یک کنل $K \times K \times J$ است، پس تعداد وزن ها در این حالت $I \times J \times K \times K$ است.

\leftarrow CNN ها پارامتر لسته دارند و به داده آموزشی لسته نیاز دارند و نتیجه را مبدل ب تراویر انتساب می کنند. (گرچه مکاره ایست موارد استثنایی داشته باشند \rightarrow این عبارت لزوماً همواره برقرار نیست).

ب) درسته هار FC و $Batch Norm$ درسته هار FC و $Batch Norm$ را در تفاوت نموده ایم

درسته هار FC درسته هار $Batch Norm$ درسته هار CNN درسته هار $Feature map$ درسته هار $Feature map$ درسته هار $Batch Norm$ درسته هار CNN

میانیز داریم $Feature map$ محاسبه شود سیستم کار سازی می شود. این کار $Batch Norm$ باعث می شود که افزونی تابع همبستگی مکانی لسته شود و بیشتر داده هار CNN بحث ایست.

بعد از $Batch Norm$ ایستاده می کرد ممکن احتلالی درسته ایجاد نمی شود

زیرا که لایر $z = \alpha + \beta \cdot \text{فیلت}$ $Batch Norm$ باشد آنها اثر نمود $bias$

صیغه می شود

حالا اگر وزن هارا در α ضرب کنیم، خواهیم داشت

$$x_{in} \xrightarrow[\text{layer}]{\text{Conv}} \alpha x_{out} \xrightarrow[\text{Norm}]{\text{Batch}} \alpha Y \left(\frac{\alpha x_{out} - \alpha \mu}{\alpha \sigma} \right) + \alpha \beta$$

$$= \alpha \left(\frac{x_{out} - \mu}{\sigma} \right) + \alpha \beta = \alpha y$$

حالا $f(y), f(\alpha y)$ مانند activation مانند linear رابطه دارند (مسوی) برای y برسی کنیم.

حالا اگر inference مانند forward pass و backward pass خواهیم داشت

$$\alpha x_{in} \xrightarrow[\text{layer}]{\text{Conv}} \alpha x_{out} \xrightarrow[\text{Norm}]{\text{Batch}} \cancel{\text{Batch Norm}} \gamma \left(\frac{\alpha x_{out} - \alpha \mu}{\alpha \sigma} \right) + \beta =$$

$$\left(\frac{x_{out} - \mu}{\sigma} \right) + \beta$$

عملکرد مثل تعمیر normalization است

در حوزه هایی مثل تعلیم پریشی که سنتی بکسل هار و ویدیو توزیع نامحولی داشت، نرمال سازی

مکانیسمی باعث شود خروجی ما ضمیمه شود - یعنی مقدار دیتاها / pixel / CNN

نیز این براحتی تعداد تولیدر دقیق نیستند عموماً یک کلر سفید و سطح پس زیر طبقه نمود

دسته های بسیار: استاده از Instance Normalization

همچنین صیغه ای از Group Normalization —

که استاده از

۲) مطابق ساختار UNet، ماجهور بارگاهن بعد دارای 14×14 Feature map، مخواهد بود.

در وحی لایر دوم 128×128 کانال دارد و خروجی این 92×92 کانال \rightarrow مرحله دریغه عمیق هر کدام است.

$$\Rightarrow 3 \times 3 \times 128 \times 64 = \text{تعداد پارامترها}$$

۳) همه سلسله ها / deconvolutional به این اکثر رایج دند که نیز هم محصور که منجر به معالجه نهاده هار خاص می شود را بازسازی کنند. این درست \downarrow بار استفاده از max-pool، از روی جیلیکر ہنام Guided-Back Prop استفاده می کند باعث اینجا دفعات بیشتر و دقیق تر می شود. در این روی لایه هار یا سین تر ویرجی هار مسأله مثل که ها را مستحب می دهند و لایه هار بالاتر \downarrow هار abstract تر مثل وابستگی به کلاس داده را فراهم می کنند \leftarrow این مدد به بصور تفسیر پذیر CNN ها کند می کنند.

سلسله هار \uparrow -Conv به بازسازی دفعاتی با استفاده از برطر هار ویرجی استخراج سه ترکیب می کند و باعث تفسیر پذیر بحث مدل \downarrow عالیه این دو شیوه upsampling و Convolution، ترکیب می کند و باعث می کند \downarrow جزئیات زیاد، مکمل، \downarrow این را بازسازی کنند و همین \downarrow نتایج می دهد ویرجی هار مکانی، جزئیات یا اطلاعات که در دو لایه قبلی ذخیره می کنند. \rightarrow بیشترین بینی سلسله و تفسیر پذیر بحث آرٹیفیشیل

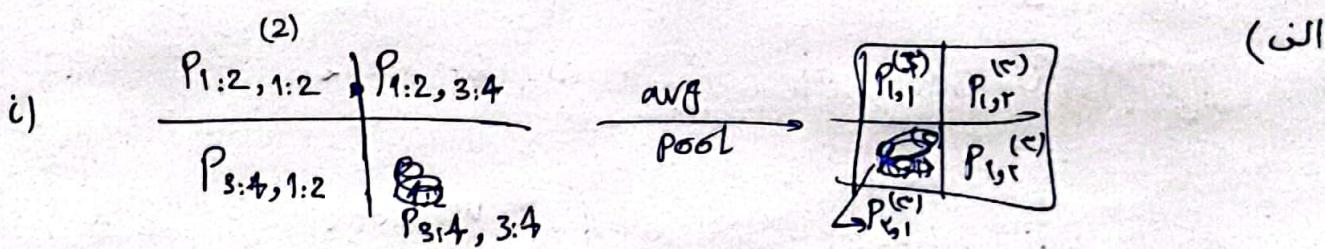
$$\frac{\partial L}{\partial w_j} = \sum_{i=0}^{N-K} \frac{\partial L}{\partial z_i} \cdot \frac{\partial z_i}{\partial w_j} = \sum_{i=0}^{N-K} \frac{\partial L}{\partial z_i} \sum_{s=0}^{K_1} \frac{\partial w_s}{\partial w_j} x_{is} = \sum_{i=0}^{N-K} \frac{\partial L}{\partial z_i} x_{ij}$$

آنچه حضرت از درایو صوت $\hat{f} = [f_1, \dots, f_{N-K}]$ در تدریل بلکه میرسید f_i را $\frac{\partial L}{\partial z_i}$ خواهد داشت

لابطر بالا متن اینجا بدلن \hat{f} را \times است و باید باشد \hat{f} بدلن \hat{f} را \times است و باید باشد

$$\frac{\partial L}{\partial w_j} = \sum_{i=0}^{N-K} f_i x_{ij}$$

$$\hat{y} = \sigma \left(w_{1,1} p_{1,1}^{(r)} + w_{1,1} p_{1,1}^{(r)} + w_{1,2} p_{1,2}^{(r)} + w_{1,2} p_{1,2}^{(r)} \right) \quad -3$$



$$\frac{\partial L}{\partial p_{i,j}} = \frac{\partial \sigma(z)}{\partial z} \frac{\partial z}{\partial p_{i,j}} = \hat{y}(1-\hat{y}) w_{i,j}$$

sigmoid مسئن

(ii)

$$p_{i,j}^{(r)} = \frac{1}{r} \sum_{k=1}^r \sum_{s=1}^s p_{r(i-1)+k, r(j-1)+s}^{(r)}$$

$$p_{i,j}^{(r)} = \sum_{k=1}^r \sum_{t=1}^t \sum_{s=1}^s w_{k,t,s}^{(r)} p_{i+k, j+t, s}^{(r)}$$

$$\frac{\partial L}{\partial p_{i,j}^{(r)}} = \frac{\partial L}{\partial p_{i,j}^{(r)}} \cdot \frac{\partial P_{i,j}^{(r)}}{\partial p_{i,j}^{(r)}} = \frac{1}{r} \hat{y}(1-\hat{y}) w_{i,j}^{(r)} \quad \text{الخطوة الأولى}$$

$$\Rightarrow \frac{\partial L}{\partial p_{i,j,k}^{(l)}} = \sum_{t=1}^T \sum_{s=1}^S \frac{\partial L}{\partial p_{t,s}^{(r)}} \frac{\partial P_{t,s}^{(r)}}{\partial p_{i,j,k}^{(l)}} =$$

$$\sum_{t=i-1}^i \sum_{s=j-1}^j \frac{\partial L}{\partial p_{t,s}^{(r)}} \frac{\partial P_{t,s}^{(r)}}{\partial p_{i,j,k}^{(l)}} = \sum_{t=i-1}^i \sum_{s=j-1}^j \frac{1}{r} (\hat{y}-1)\hat{y} w_{i,j}^{(r)} \frac{\partial L}{\partial \hat{y}}.$$

$w_{i-t+1, j-s+1}^{(2)}$

$$P_{i,j,m}^{(l)} = \sum_{s=1}^S w_{i,j,s}^{(m)} P_{i,j,s}^{(r)} \quad \text{المتغيرات المدخلة في المراحل المبكرة}$$

$$\frac{\partial L}{\partial w_{i,j,k}^{(m)}} = \sum_{i=1}^T \sum_{j=1}^S \frac{\partial L}{\partial p_{i,j,m}^{(l)}} \cdot \frac{\partial p_{i,j,m}^{(l)}}{\partial w_{i,j,k}^{(m)}} = \sum_{i=1}^T \sum_{j=1}^S \frac{\partial L}{\partial p_{i,j,m}^{(l)}} \cdot P_{i,j,m}^{(l)}$$

$$\frac{\partial L}{\partial w_{i,j,k}^{(l)}} = \sum_{t=1}^T \sum_{s=1}^S \frac{\partial L}{\partial p_{t,s}^{(r)}} \frac{\partial P_{t,s}^{(r)}}{\partial w_{i,j,k}^{(l)}} = \sum_{t=1}^T \sum_{s=1}^S \frac{\partial L}{\partial p_{t,s}^{(r)}} P_{t+s, i+j, k}^{(l)} \quad (2)$$

بعض القيم على حساب الخطأ

- ۱۵ ResNet خود x را هم علاوه بر $(x)F = \nabla F$ بعدی می‌دهیم یعنی خروجی دستالیه

با خروجی دستالیه قبلی لست جمع مُدّه است. این کار باعث می‌شود که از مشکل vanishing gradient

جلوگیر کنیم، زیرا نهادیان علک ارس تا کوچک شود دلیل افتد که خروجی دستالیه قبلی از این کار جلوگیر می‌کنیم.

DenseNet خروجی تمامی دستالیه‌ها قبلی را به درودی فعلی می‌دهیم، باعث می‌شود که از خوبی هایی که در دستالیه‌ها قبلی یاد نموده ایم نیز استفاده کنیم. انتقال هر از طریق concat می‌شود.

تفاوت ها، ~~ResNet~~

انتقال از طریق جمع ارسی ResNet با دستالیه قبلی فقط یعنی کافی است / حافظه کشیده و مصروف نیست.

~~DenseNet~~ انتقال از طریق ~~concat~~ ارسی است. DenseNet تمام دستالیه‌ها / قبلی concat شوند / حافظه بسیار مصروف نیست.

DenseNet بدل انتقالات می‌نماید. ResNet رفع می‌شود. زیرا هر دوی به دلیل مستقر پیشام دستالیه $\text{vanishing gradient}$.

قبلی دستالیه دستالیه‌های پیش از خود مستقر بودند این دستالیه‌ای باز نباشد، این باعث کاهش طبقه می‌شود که دستالیه نیزی نمود. بین صورت با کاهش عمق دستالیه بدلیل نهادیان نایاب شوند

در ~~ResNet~~:

در سرالیه عکس کمال جنس لغایه می‌گذارد

$$\rightarrow \text{input-channels} + k \times 2^t = 32 + 3 \times 2^4 = 104$$

(—)

- باز از هر نورده هزوی در حالت اول پی علیات با کرنل اجم محدود هر علیات

operating system $5 \times 5 \times 192$

$5 \times 5 \times 192 \times 28 \times 28 \times 22$

حالت دوم \rightarrow
بار کانولوئن 1×1 اجم داده ایم و ساز آن $28 \times 28 \times 192$

$28 \times 28 \times 192$ بار کانولوئن 5×5 اجم داده ایم \leftarrow

$1 \times 1 \times 192 \times 28 \times 28 \times 192 + 5 \times 5 \times 192 \times 28 \times 28 \times 22$

حالت دیگر \rightarrow

$$\frac{\text{تعداد علیات دوم}}{\text{تعداد علیات اول}} \approx 0.1$$

استفاده کنند بخطه هزوی \rightarrow moduler Inception v3 backbone google net / سینه

نسبت بیشتر نیست بلکه کار مقام ایس. در عال قبله هزوی

در شبکه ها عین وقی داریم به لذیح اولیه برخی کردیم کارایی عملکرد ایس

بسیار کمودی محو سوی در شبکه google بخطه سیمانهم

خطه با جزئی مسیر \rightarrow کارایی اجازه می دهد از جزئی بازدید و اثر کمی

آنرا \rightarrow انتشار. چنین استفاده شبکه 1×1 کارایی ندارد

کار ها و افرادی که جریان کارایی کنند چنین چنین پیشنهاد کنند

و قریب در تمام لایه شبکه کم می کنند

که google net / plasitc را دوستی داشت که به کمک Auxilliary classifier

باید شرط کارایی تک حالت classifier در الایتم های مجموعه فکری و خود را در آن classifier میگشتند. این classifier های مخصوصاً در الایتم های مجموعه فکری و خود را در آن classifier میگشتند.

استخراج کنند. این کار باعث شد دوستی دوستی داشته باشند.

با اینکه نکات بسیاری از این classifier های مخصوصاً در الایتم های مجموعه فکری و خود را در آن classifier میگشتند، با اینکه نکات بسیاری از این classifier های مخصوصاً در الایتم های مجموعه فکری و خود را در آن classifier میگشتند.

محدودیت های این طبقه عبارتند از اینکه: این طبقه باعث افزایش حجم مجموعه میگشند

و بدل داده های خالی بسیاری عمل نمیکنند. جزئیه نسبت زیرا که این سروی های مخصوصاً

محدودیت های این طبقه عبارتند از استخراج نکرد.

(2) در شرطی دسته های ~~convolution~~ عادی:

سایر فناوری های این است. ~~offset~~ پیشنهاد میکند مطالعه (stride) و پیدنده مفهومی (geometric) نیز فناوری های این است.

در شرطی دسته های ~~convolution~~ عادی:

متوجه شد ~~offset~~ و ~~offset~~ تغییر مکان (offset) را در

جهار یادگیر نهاده نماید در نفعی خواهد بود. مطالعه باشید /
 ~~offset~~ کافته میگشند.

برای درست این اتفاق بتوانیم بتوانیم در عبارت $\text{conv}_k^{\text{in}} \rightarrow \text{conv}_k^{\text{out}}$ این k را بسیار کم

دان.

(deformable curves, grid-sampling)

در کادو ایندیکاتور مانند، منونه بر طریق ارزش تابع $\text{Grid}_{\text{نحوی}}(\cdot)$ ایجاد می‌گردد.

متلا بکر نیلیت ۳۴، نگاه مکرر انتقامی از

در سنبله deformable) این بند مستقر تغییر می‌کند و نتایج نمونه برآورد پهلوک ها را جیز (غیر مراقب) می‌نماید.

offset بیتکه ای دلخواه
net نسبتی که از مجموع
مقدارها کمتر است

(GT بھرپور سے ہے)

جول مکان نتایج نمونه بخاری (deformable) ہے، تا خوبی اس کو لکھوں۔

مختصر - راهنمای ایران سخنی دیدار

• میں اپنے offset کو نئی لینڈ کارڈ پر کامیاب کر سکتے ہیں.

مکالمہ جسی نکاط بطر آفسنگ = اکٹھ لے کر نکالو

- ۵

الن) مدل $F \times D$ با $M - D(F-1)-1 + 1 = D$ برمی عکل دید سطح حرکت
در $N - D(F-1)-1 + 1$ بار حرکت می کوئن حرکت داد

$$(M - DF + D) \times (N - DF + D) = \text{سایز خروجی}$$

 زیرا داستانه: زیرا ~~ذرا~~ ~~ذرا~~

$$\text{output-size} = \left\lfloor \frac{M - D(F-1)-1}{\text{stride}} \right\rfloor + 1$$

معامله کننده k' بعد از $D-F+1$ نیز است،

$$k' = \begin{bmatrix} k_{11} & \dots & k_{12} & \dots & k_{1F} \\ \vdots & & \diagdown & & | \\ 1 & & & & | \\ \vdots & & & & | \\ k_{F1}^0 & \dots & \dots & \dots & k_{FF} \end{bmatrix}$$

طبق ضرب در فندر:

$$K \otimes A = \begin{bmatrix} k_{11}A & k_{12}A & \dots & k_{1F}A \\ | & | & & | \\ k_{F1}A & \dots & \dots & k_{FF}A \end{bmatrix}$$

برای این $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ مدل k' بسته می آید، معین $D-1$ سطح سطر آخر
صفراند که جمله تأثیر در conv_1 می کلن آنرا نادیده نمود.

(2)

جی کا میکسٹر ملائیں،

اگر میکسٹر نہیں تو جی میکسٹر

$I_r(i, j)$ \rightarrow آئے در I_r عبارت اسے رہے receptive field

$$I_r(i, j) \rightarrow I_r(i + (d+1)w, j + (d+1)w)$$

اگر I_r ہر حصہ کا d راتلکاری کرنے والا

$$I_r(i', j') \rightarrow \begin{cases} I_r(i', j') \\ I_r(i' + d(w-1), j' + d(w-1)) \end{cases}$$

اگر I_r ہر حصہ کا d راتلکاری کرنے والا

$$I_r(i'', j'') \rightarrow I_r(i'', j'') - I_r(i'' + (d-1)(w-1), j'' + (d-1)(w-1))$$

$$\rightarrow I_r(i, j) = I_r(i + (d+1)w + d(w-1) + (d-1)(w-1), \\ j + (d+1)w + d(w-1) + (d-1)(w-1))$$

mask mask، نوعی کانولوشن صورت میں خروجی طور پر masked conv ()

امالی mask کانولوشن روکھوں، معنی کوتا مل بازی لفڑا دیر ہیا دل

اسکالپریوگر، بیلر لینہ مدل معوانہ حدس بزندہ، Pixel CNN

کے خانہ را بچھتا یقینہ نہ کر سکتا۔ (وہی کہ نبی خواصیں ملے بچھی از طرف کسی

درستہ بارے)۔

کاربرد Masked conv محدوده مدل سازی خود را تحریکی و پردازشی خود را باقی استفاده می کند

مثل Pixel CNN. این ماسک باعث می کند که نول در تغابه اطلاعات قبلی در توالي دسترسی داشته باشد و زده های آنها بزرگتر بینی استفاده نکنند. در محدوده مدل ها باعث نمایش به شکل این اجزای را می دهد تا دفعه های بازسازی کند.

محدوده های این مدل هر عبارت را

باعث بسیاری مطالعه می کند و میتواند مطالعه را حذف کند. این محدوده دقت مدل را تکثیر کند. نیاز به درک عمیق از اختصار داده دهنده محدوده می خواهد زیاد کند.

آن دلخواه نسخه ای از محدوده هار Masked conv (ساختاری که نشان دهد. این نوعی

با قریبی عاملی سیم می بردی) محدوده را در conv را برای افرادی که با اینها بزرگ شده اند. با اینکه اطلاعات سیم از سیم باشند می باشد. در حالی که محدوده های ماسک چنانچه رعایت می کنند. در بازسازی تغایر یافته های مولده اطلاعات سیم را استخراج می کند و گزینی خود را با یک عدد می بینند.

$$\|w_i\|_2 = \sigma_{\max}(w_i), \quad \text{دارای} \rightarrow$$

$$\|\delta_i\|_2 = \|s_i\|_2 \prod_{k=i+1}^L \|w_k f'_{k-1}(z_{k-1})\|_2$$

$$\leq \|s_i\|_2 \prod_{k=i+1}^L \|w_k f'_{k-1}(z_{k-1})\|_2 \quad \begin{array}{l} \text{ما ترسی از مقدار بزرگ} \\ \text{میگیریم، ReLU} \\ \text{منطبق} \end{array}$$

$$\leq M \prod_{k=i+1}^L \|w_k\|_2 \|f'_k(z_{k-1})\|_2 \quad \text{(به)}$$

$$\leq M \prod_{k=i+1}^L \sigma_{\max}(w_k) \sigma_{\max}(f'_{k-1}(z_{k-1}))$$

استدلال f' (نهایت درست است)

$$\leq M \prod_{k=i+1}^L \sigma_{\max}(w_k)$$

استدلال f' (نهایت صحت دارد) $\rightarrow \sigma \leftarrow \sigma_{\max}(w_k) < 1$

$$\|\delta_i\|_2 \rightarrow 0 \quad \text{when } L \rightarrow \infty$$

$$\sigma_{\max} = \sqrt{\sigma_{\min}}$$

$$\|\delta_i\|_2 \leq M \prod_{i=1}^L \sigma = M \cdot \sigma^L \approx M \left(\frac{1}{10} \times 10^{-2} \right)$$

۱) ملحوظات:

$$w \sim N(0, \frac{\gamma}{n_{in}})$$

تعديد دردسي بيله هر نورده (افتخار)

$$n_{in} = 10 \times 10 \times 3 = 192$$

$$w \sim N(0, \frac{\gamma}{\frac{1}{192}})$$

$\frac{1}{192}$

۲) استفاده از ReLU برای sigmoids به جلوگیری از مسلسل بازگشتن نتایج

جهد هسته مان ایو اس. منتهی بیل sigmoid در نهاد بزرگ و لوحی
کفریاً صفریاً بود. معین موضوع در AlexNet باعث موفقیت شد.

۳) وقتی که دردسي بيله نورده همراه هسته صفری بود، علاوه

آن نورده تغییر نمی کند (dead neurons) وزنها را

آبیسی نمی کردند

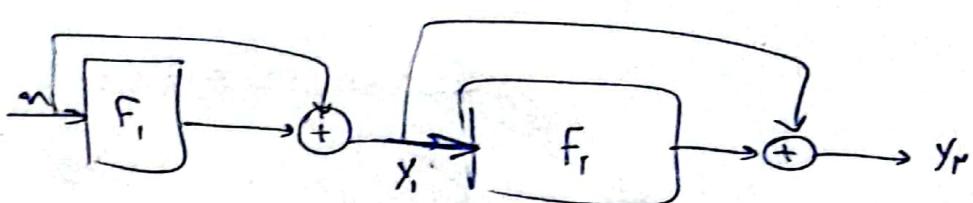
---, ELU, PReLU, leaky-ReLU, ~~جيكتور~~ ^{جيكتور} را همان

introduce

قدیمات ایج توانع فعالیت گرم اس لایرها آمده اند.

حال آنچه بررسی می کنیم

ReLU	صائب و نووو مرد	منیت ساده ارس
Leaky ReLU	متا دی رسمی پذیر نی خواهد	نو رو زه مرد هزار بیز
PReLU	بی خوبی بینه	انعطاف پذیر باز
ELU	صائب سه سلکتر	منیت پس شود متا دی رسمی



(8)

$$y_1 = F_1(x) + x$$

$$y_2 = F_r(y_1) + y_1$$

حال نکارا بروزی می کنیم

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial y_r} \frac{\partial y_r}{\partial x} = \frac{\partial L}{\partial y_r} \left(\frac{\partial F_r}{\partial x} + \frac{\partial y_1}{\partial x} \right)$$

$$= \frac{\partial L}{\partial y_r} \left(\frac{\partial F_r}{\partial y_1} \underbrace{\frac{\partial y_1}{\partial x}}_{+1} + \frac{\partial y_1}{\partial x} \right) =$$

$$\frac{\partial L}{\partial y_r} \left(\frac{\partial F_r}{\partial y_1} + 1 \right) \left(\frac{\partial y_1}{\partial x} \right) = \cancel{\frac{\partial L}{\partial y_r} \left(\frac{\partial F_r}{\partial y_1} + 1 \right)} \cancel{\left(\frac{\partial y_1}{\partial x} \right)}$$

$$\frac{\partial L}{\partial y_r} \left(\frac{\partial F_r}{\partial x} + 1 \right) \left(\frac{\partial F_1}{\partial x} + \frac{\partial x}{\partial x} \right)$$

ناریم skip connection

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial y_i} \frac{\partial F_i}{\partial y_i} \frac{\partial F_i}{\partial w}$$

استفاده از residual block
با عناصری که در ماتریس تریده باشند، باعث دارم در
سرو آن میتواند بنا برای آنکار به صفت جیل از طبقه کند میکند.

~~دستوراتی که در ماتریس تریده باشند~~

~~دستوراتی که در ماتریس تریده باشند~~

دستوراتی که در ماتریس تریده باشند

اعمال پس از فرآیند که در وجود دارد