



2019.10.10

Siam 网络加深探究 SiamRPN++ & siamDW

CVPR19_oral

这两篇论文都对 siam 网络的加深问题进行了探究，分析了限制加深的原因，以及相应的解决办法。两篇论文都是 CVPR2019oral，看来 siam 的深度问题确实很受关注啊。

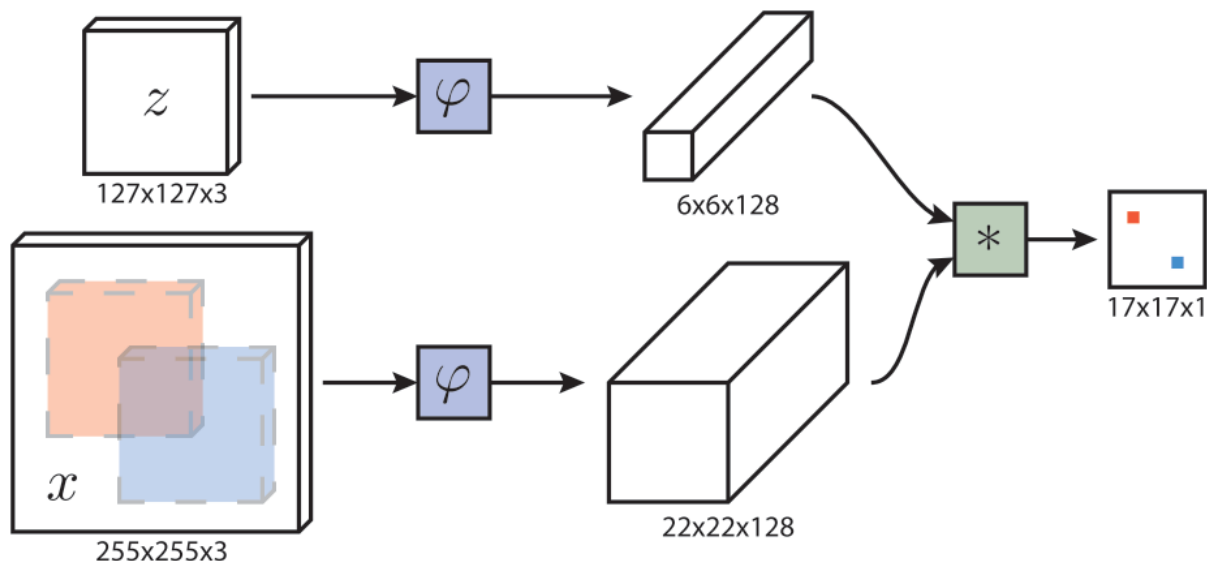


siamRPN++ 这篇论文认为导致 siam 系列网络不能加深的阻碍在于加深网络必然要引入 padding，但是按照 siamFC 的训练方式，如果加入 padding 会引入 spatial bias，导致虽然网络深度上去了，但是跟踪的效果却反而下降。

因此 siamRPN++ 的解决方式是从训练方式上解决 -> 在训练时引入偏移，消除 spatial bias。

此外，siamRPN++ 除了映射网络加深了之外，还融合了 FPN 特征金字塔的思想，进一步提高准确度，随之带来的是参数量的上升，为了减少参数，又引入了 mobile net 的思想，采用 depthwise 卷积。

siamFC 网络



Siamese network 速度很快，但是准确度不够， ϕ 的深度不够是主要原因。



不能 padding，导致加深网络。siamfc 的作者说会 padding 会改变 translation equivariance. **siamRPN++ 的作者认为引入 padding 会引入 spatial bias。**



进行实验证明作者猜想正确。

证明 padding 会引入 spatial bias

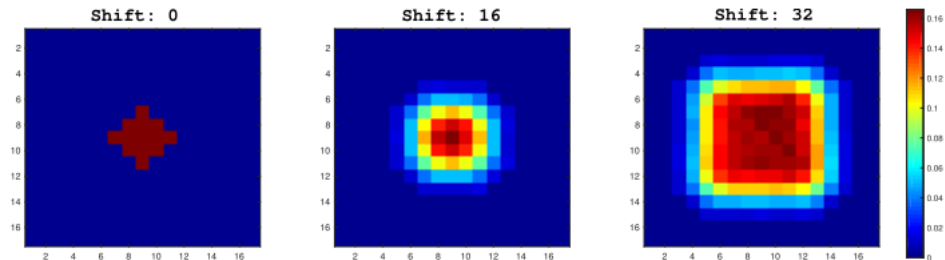
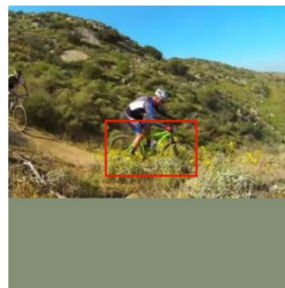


Figure 1. Visualization of prior probabilities of positive samples when using different random translations. The distributions become more uniform after random translations within ± 32 pixels.



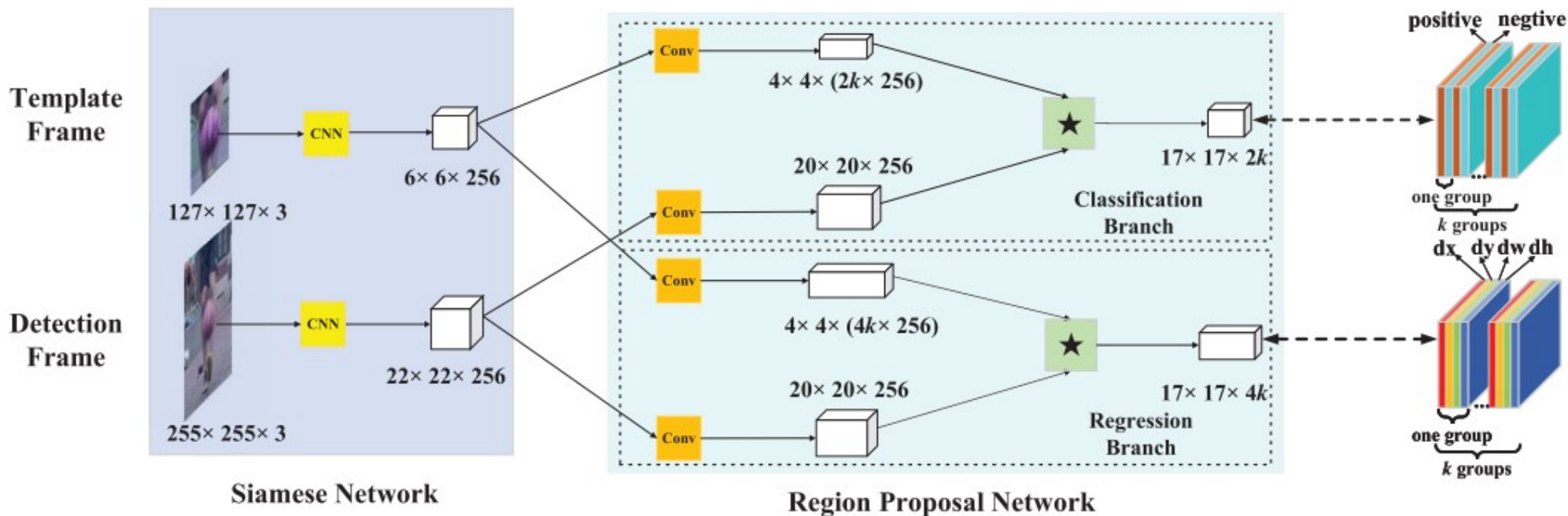
Figure 2. The impacts of the random translation on VOT dataset.



证明了在原来的训练方法下，使用 padding 确实会引入 spatial bias。网络偏向认为图片中心处就是目标。且均匀采样有效。



证明了让 target 产生均匀偏移，对于跟踪性能有提高。



regression :

$$\delta[0] = \frac{T_x - A_x}{A_w}, \quad \delta[1] = \frac{T_y - A_y}{A_h}$$

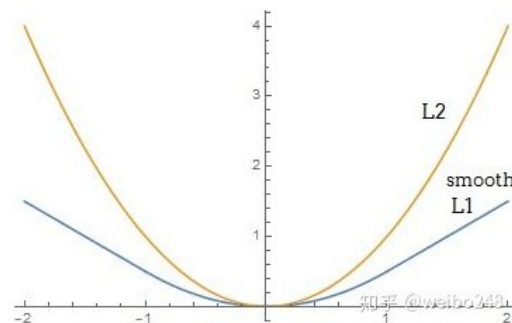
$$\delta[2] = \ln \frac{T_w}{A_w}, \quad \delta[3] = \ln \frac{T_h}{A_h}$$

cls → cross-entropy loss; reg → smooth L₁ loss

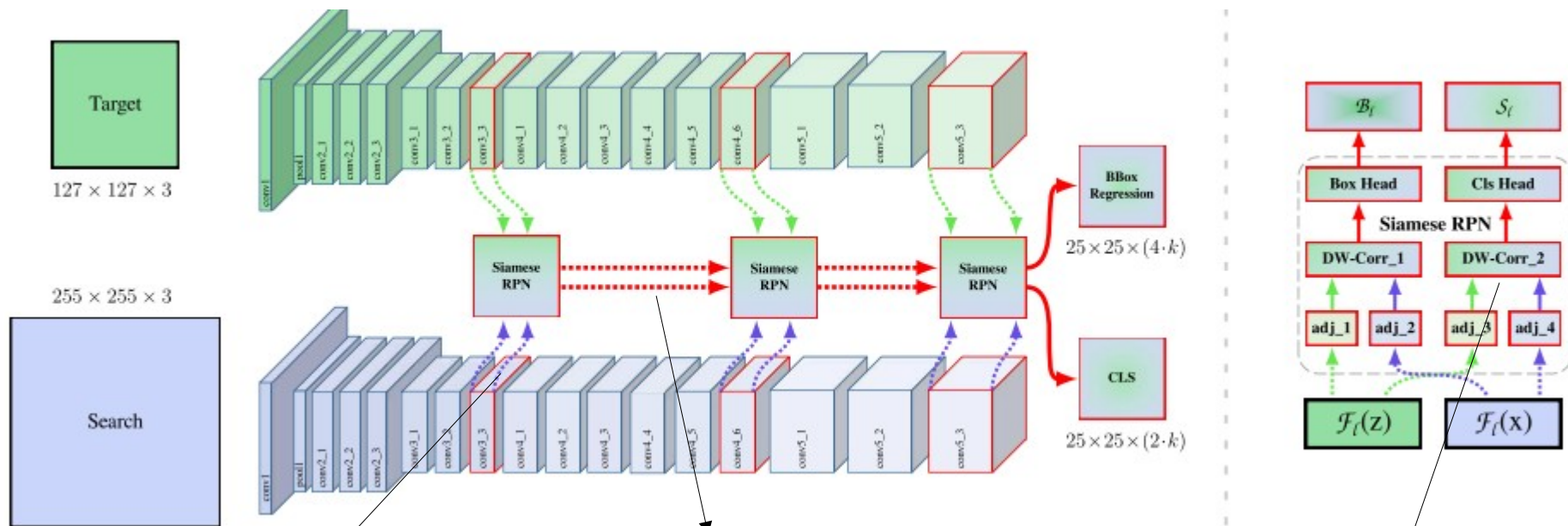
$$\text{loss} = L_{\text{cls}} + \lambda L_{\text{reg}}$$

L₁ loss:

$$\text{smooth}_{L_1}(x, \sigma) = \begin{cases} 0.5\sigma^2 x^2, & |x| < \frac{1}{\sigma^2} \\ |x| - \frac{1}{2\sigma^2}, & |x| \geq \frac{1}{\sigma^2} \end{cases}$$



相比于 L₂, L₁ 对于离群点更加鲁棒。



layer-wise aggregation,
综合各类特征。

将各自的输出分别相加。

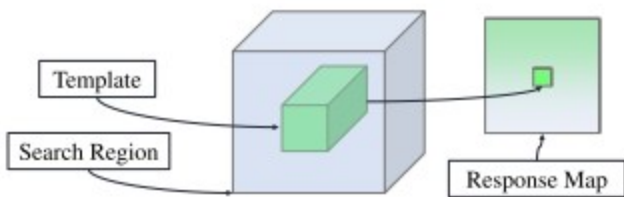
$$S_{all} = \sum_{l=3}^5 \alpha_l * S_l, \quad B_{all} = \sum_{l=3}^5 \beta_l * B_l.$$

不仅是深度网络，还融合了浅网络。类似于 FPN 网络的做法特征融合。

Depthwise cross correlation,
将原来 siamRPN 在 correlation 环节的
参数减少了至 0.1 倍。
The RPN module contains 20M parameters
while the feature extractor only contains 4M
parameters in

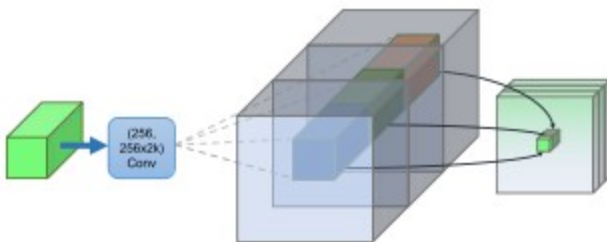
Depthwise cross correlation

siamFC



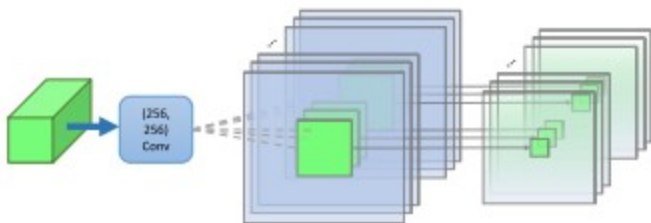
(a) Cross Correlation Layer

siamRPN

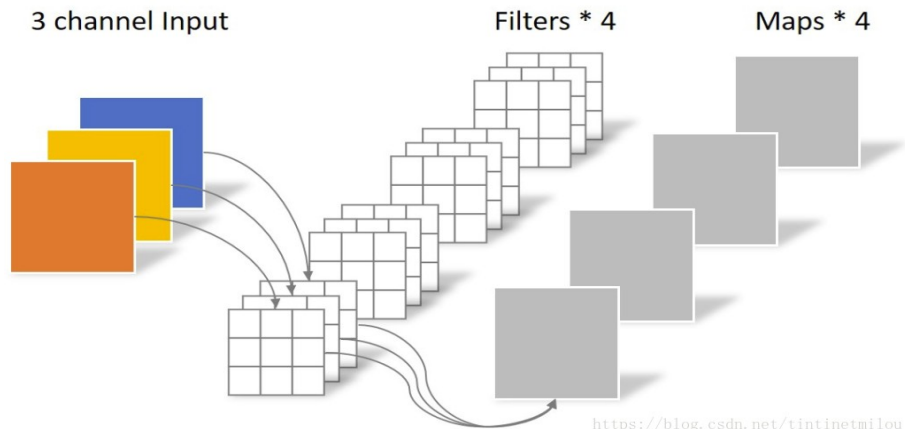


(b) Up-Channel Cross Correlation Layer

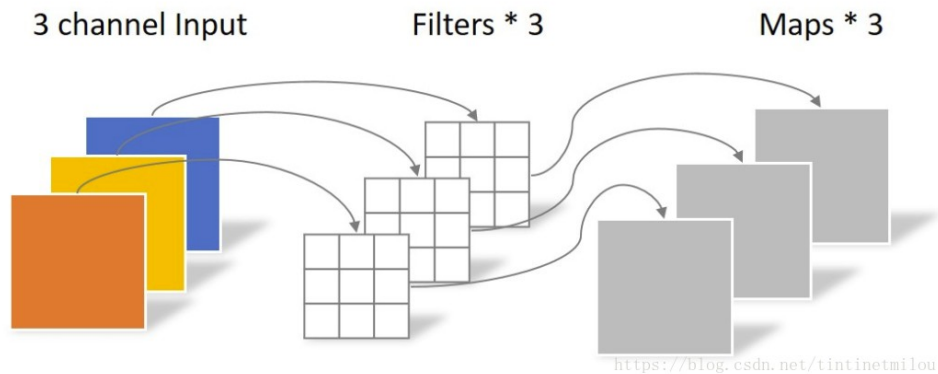
siamRPN++



(c) Depth-wise Cross Correlation Layer

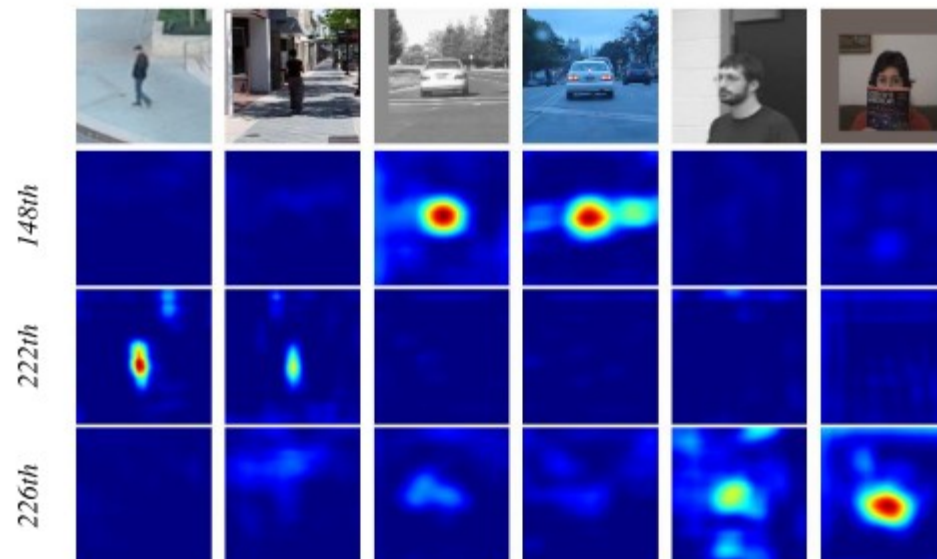
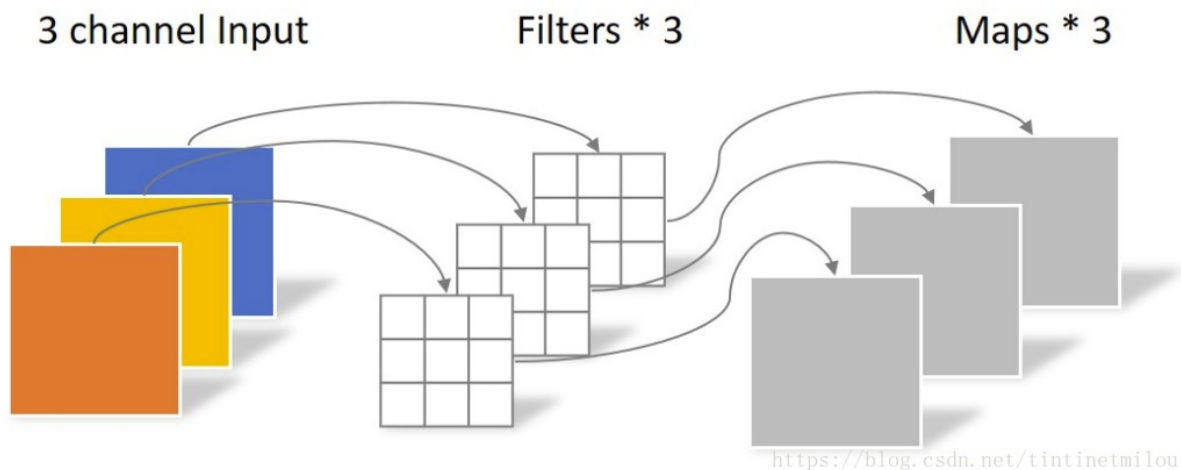


Pointwise convolution



Depthwise convolution, 计算量减低。

Interesting phenomena about depthwise



同一类物体的最大响应值出现在同一张 feature 上，而其他 feature map 上则响应很小。作者认为深度网络提取到 feature map 是正交的。 -> 减少深度网络参数的一种方法？



	DLSTpp	DaSiamRPN	SA_Siam_R	CPT	DeepSTRCF	DRT	RCO	UPDT	SiamRPN	MFT	LADCF	Ours
EAO \uparrow	0.325	0.326	0.337	0.339	0.345	0.356	0.376	0.378	0.383	0.385	0.389	0.414
Accuracy \uparrow	0.543	0.569	0.566	0.506	0.523	0.519	0.507	0.536	0.586	0.505	0.503	0.600
Robustness \downarrow	0.224	0.337	0.258	0.239	0.215	0.201	0.155	0.184	0.276	0.140	0.159	0.234
AO \uparrow	0.495	0.398	0.429	0.379	0.436	0.426	0.384	0.454	0.472	0.393	0.421	0.498

Table 2. Comparison with the state-of-the-art in terms of expected average overlap (EAO), robustness, and accuracy on the VOT2018.

- 对 siamFC 类进行改进，分析其不能使用深度网络的原因，即会引入 spatial bias 。并通过实验证明，同时对 target 进行偏移可以解决此问题，使得网络可以加深。
- 基于上述探究，对 siamRPN 进行改进，融合了网络不同层的特征，提高准确度，达到 sota ，为了降低计算开销，使用 depthwise convolution 。



siamDW 认为 siam 系列网络不能加深原因有二：第一，随着网络深度的增加，网络的感受野增加，从而减少了网络的判别性与回归的准确度；第二，padding 会引入 spatial bias，因为如果使用 padding 的话，对于卷积核（template）来说是一定带 padding，而对于 search image 中间部分是没有 padding 的，只有边缘的才有 padding，作者认为这会导致不连续，导致对于 search image 边缘的目标识别很差。

因此本文的作者从两个方面探究了加深 siam 系列网络的办法：第一，感受野的问题，作者探究发现 siam 系列网络 prefer small stride, 4~8 最宜，同时网络最后的感受野最好在整幅 exemplar 的 60% 至 80% 最佳，stride 也要根据这个来调整；针对 padding 问题，作者设计了一种新的参差网络结构，先 padding，再删除受 padding 影响神经元（不太理解这的作用，仿佛和 no-padding 没啥差别）。

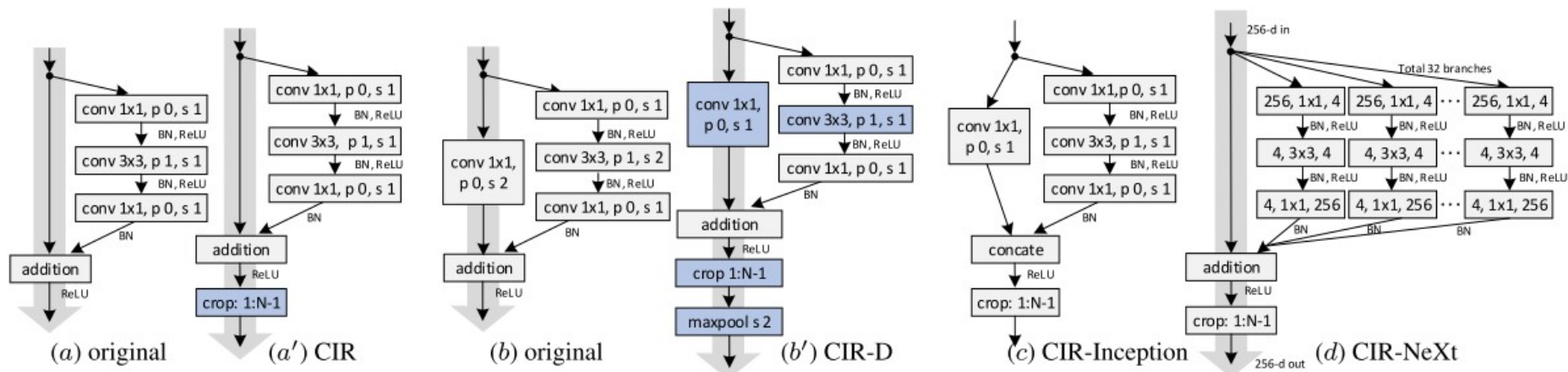


Figure 3: The proposed cropping-inside residual units. (a) and (b) are the original residual unit and downsampling unit, while (a') and (b') are our proposed ones. (c) and (d) are the proposed wide residual units. The grey arrows indicate the shortcut paths for easy information propagation, while the blue boxes highlight the differences from the original units. The letters 'p' and 's' indicate the padding size and stride, respectively. The settings of 'p' and 's' in (d) are the same as in (c).

作者的思路是既然 padding 会产生负面影响，但是不 padding 网络又没法加深。那不如在进行卷积之前先 padding，保证 feature map 的宽度能够正常卷积，在卷积结束后把受到 padding 影响的最外层删除掉。不就使得既能正常加深，又能不引入 padding 的负面影响。

这和 no padding 似乎没啥区别？？？



Table 5: Performance comparisons on five tracking benchmarks. **Red**, **Green** and **Blue** fonts indicate the top-3 trackers, respectively.

Tracker	Year	OTB-2013		OTB-2015		VOT15			VOT16			VOT17		
		AUC	Prec.	AUC	Prec.	A	R	EAO	A	R	EAO	A	R	EAO
SRDCF [5]	2015	0.63	0.84	0.60	0.80	0.56	1.24	0.29	0.54	0.42	0.25	0.49	0.97	0.12
SINT [34]	2016	0.64	0.85	-	-	-	-	-	-	-	-	-	-	-
Staple [1]	2016	0.60	0.80	0.58	0.78	0.57	1.39	0.30	0.54	0.38	0.30	0.52	0.69	0.17
SiamFC [2]	2016	0.61	0.81	0.58	0.77	0.53	0.88	0.29	0.53	0.46	0.24	0.50	0.59	0.19
ECO-HC [4]	2017	0.65	0.87	0.64	0.86	-	-	-	0.54	0.3	0.32	0.49	0.44	0.24
PTAV [8]	2017	0.66	0.89	0.64	0.85	-	-	-	-	-	-	-	-	-
DSiam [12]	2017	0.64	0.81	-	-	-	-	-	-	-	-	-	-	-
CFNet [35]	2017	0.61	0.80	0.59	0.78	-	-	-	-	-	-	-	-	-
StructSiam [40]	2018	0.64	0.88	0.62	0.85	-	-	-	-	-	0.26	-	-	-
TriSiam [7]	2018	0.62	0.82	0.59	0.78	-	-	-	-	-	-	-	-	0.20
SiamRPN [20]	2018	-	-	0.64	0.85	0.58	1.13	0.35	0.56	0.26	0.34	0.49	0.46	0.24
SiamFC+	Ours	0.67	0.88	0.64	0.85	0.57	-	0.31	0.54	0.38	0.30	0.50	0.49	0.23
SiamRPN+	Ours	0.67	0.92	0.67	0.90	0.59	-	0.38	0.58	0.24	0.37	0.52	0.41	0.30

