Table 2: Performance (accuracy, %) contributions of the components in RBCNs on CIFAR100, where Bi, R, G, and B denote the Bi-Real Net, $RBConv$, GAN, update of the BN layers, respectively. The bold numbers represent the best results.

	Kernel Stage	Bi	R	R+G	R+G+B
RBCN	32-32-64-128	54.92	56.54	59.13	**61.64**
RBCN	32-64-128-256	63.11	63.49	64.93	**65.38**
RBCN	64-64-128-256	63.81	64.13	65.02	**66.27**

Table 3: Classification accuracy (%) based on ResNet18 and WRN40 on CIFAR10/100. The bold represent the best results among the binary networks.

Model	Kernel Stage	Dataset	
		CIFAR -10	CIFAR -100
ResNet18	32-32-64-128	92.67	67.07
ResNet18	32-64-128-256	93.88	72.51
ResNet18	64-64-128-256	94.57	72.89
RBCN (ResNet18)	32-32-64-128	89.03	**61.09**
RBCN (ResNet18)	32-64-128-256	90.67	**65.38**
RBCN (ResNet18)	64-64-128-256	90.40	**66.27**
WRN22	64-64-128-256	95.19	76.38
WRN40	64-64-128-256	94.92	74.91
RBCN (WRN22)	64-64-128-256	93.28	**72.06**
RBCN (WRN40)	64-64-128-256	93.69	**73.08**
XNOR (ResNet18)	32-32-64-128	71.01	43.58
XNOR (WRN22)	64-64-128-256	86.90	58.05
Bi-Real (ResNet18)	32-32-64-128	85.34	54.92
Bi-Real (WRN22)	64-64-128-256	90.65	68.51
PCNN (ResNet18)	32-32-64-128	85.50	55.66
PCNN (WRN22)	64-64-128-256	91.62	70.32
Scheme-A (ResNet18)	32-64-128-256	75.45	46.32
Scheme-A (WRN22)	64-64-128-256	87.83	59.54

sults. As shown in the R column in Table 2, RBCN achieves 1.62% accuracy improvement over Bi-Real Net (56.54% vs. 54.92%) using the same network structure as in ResNet18 with 32-32-64-128. This significant improvement verifies the effectiveness of the learnable matrixs.

2) In RBCNs, if we use the GAN to help binarization, we can find a more significant improvement from 56.54% to 59.13% with the kernel stage of 32-32-64-128, which shows that the GAN can really enhance the binarized networks.

3) We find that a training trick can also improve RBCNs, which is to update the BN layers with W and C fixed after each epoch (line 17 in Algorithm 1). This trick makes RBCN boost 2.51% (61.64% vs. 59.13%) in CIFAR100 with 32-32-64-128.

3.3 Accuracy Comparison with State-of-the-Art

CIFAR10/100: The same parameter settings are used in RBCNs on both CIFAR10 and CIFAR100. We first compare our RBCNs with the original ResNet18 with different stage kernels, followed by a comparison with the original WRNs with the initial channel dimension 64 in Table 3. Thanks to the rectified process, our results on both the datasets are close to the full-precision networks ResNe18 and WRN40. Then, we compare our results with other state-of-the-arts such as Bi-Real Net [Liu *et al.*, 2018], PCNN [Gu *et al.*, 2019], Scheme-A [Mishra and Marr, 2017] and XNOR [Rastegari *et al.*, 2016]. All these BCNNs have both binary filters and binary activations. It is observed that at most 6.17% ($=$ 61.09%$-$54.92%) accuracy improvement is gained with our RBCN, and in other cases, larger margins are achieved.

ImageNet: Five state-of-the-art methods on ImageNet are chosen for comparison: Bi-Real Net [Liu *et al.*, 2018], BinaryNet [Courbariaux *et al.*, 2016], XNOR [Rastegari *et al.*, 2016], PCNN [Gu *et al.*, 2019] and ABC-Net [Lin *et al.*, 2017]. Again, these networks are representative methods of binarizing both network weights and activations and achieve state-of-the-art results. All the methods in Table 4 perform the binarization of ResNet18. The results in Table 4 are quoted directly from their papers, except that the result of BinaryNet is from [Lin *et al.*, 2017]. The comparison clearly indicates that the proposed RBCN outperforms the five binary networks by a considerable margin in terms of both the top-1 and top-5 accuracies. Specifically, for top-1 accuracy, RBCN outperforms BinaryNet and ABC-Net with a gap over 16%, achieves 7.9% improvement over XNOR, 3.1% over the very recent Bi-Real Net, and 2.2% over the latest PCNN. In Fig. 2, we plot the training and testing loss curves of XNOR and RBCN. It clearly shows that using our rectified process, RBCN converges faster than XNOR.

3.4 Experiments on object tracking

The key message conveyed in the proposed method is that although the conventional binary training method has a limited model capability, the proposed rectified process can lead to a powerful model. In this section, we show that this framework can also be used in object tracking. In particular, we consider the problem of tracking an arbitrary object in videos, where the object is identified solely by a rectangle in the first frame. For object tracking, it is necessary to update the weights of the network online, severely compromising the speed of the system. To directly apply the proposed framework to this application, we can construct a binary convolution with the same structure to reduce the convolution time. In this way, RBCN can be used to binarize the network further to guarantee the tracking performance.

In this paper, we use SiamFC Network as the backbone for object tracking. We binarize SiamFC as Rectified Binary Convolutional SiamFC Network (RB-SF). We evaluate RB-SF on four datasets, GOT-10K [Huang *et al.*, 2018], OTB50 [Wu *et al.*, 2013], OTB100 [Wu *et al.*, 2015], and UAV123 [Mueller *et al.*, 2016], using accuracy occupy (AO) and success rate (SR). The results are shown in Table 5. Intriguingly, our framework achieves about 7% AO improvement over XNOR, both using the same network architecture as in SiamFC Network on GOT-10k. Further, our framework brings so much benefit that Bi-SF performs almost as well as the full-precision SiamFC Network.