Visual Domain Bridge(VDB) and TENT

Pimi Yvan

Why this?

Batch normalization is a technique that stabilize and make fast the training process, but when we are in a context of different domain, and we have a source domain and a target domain, some issue can come while using batch normalization. Many approach are trying to fix it, here we are going to describe an approach which transfert the statistic of the target domain to the source domain, this approach is named Visual Domain Bridge.

Batch Normalization explained

During training time

$$\mathrm{BN}(h_c^s) = \gamma^s \times \frac{h_c^s - \mu_c^s}{\sqrt{\sigma_c^{2s} + \epsilon}} + \beta^s$$

$$\mu_c^s = \frac{1}{NHW} \sum_{n,h,w} h_{nchw}^s$$

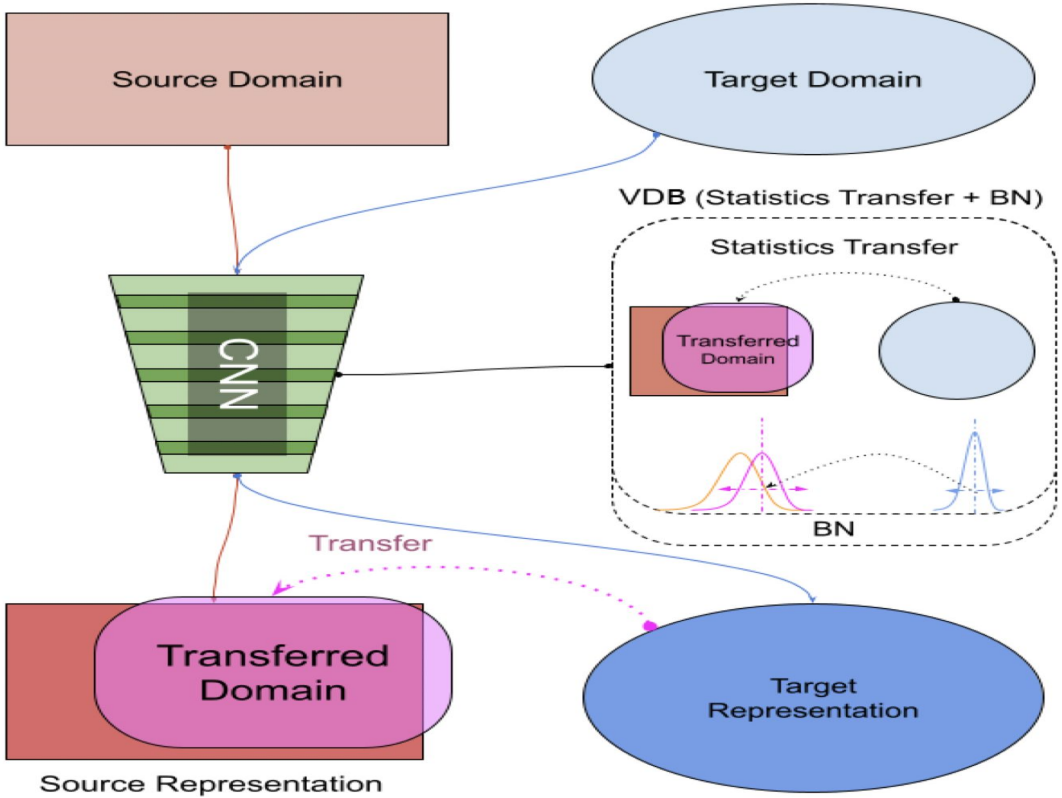
$$\sigma_c^s = \sqrt{\frac{1}{NHW} \sum_{n,h,w} (h_{nchw}^s - \mu_c^s)^2},$$

Why ?

$$\mathrm{BN}(h_c^s) = \gamma^s \times \frac{h_c^s - \mu_c^s}{\sqrt{\sigma_c^{2^s} + \epsilon}} + \beta^s$$

$$\mathrm{BN}(h_c^t) = \gamma^s \times \frac{h_c^t - \mu_c^s}{\sqrt{\sigma_c^{2^s} + \epsilon}} + \beta^s$$

Visual Domain Bridge



Visual Domain Bridge

$$\mathrm{BN}_{\mathrm{transfer}}(h_c^t) = \gamma^s \times \frac{h_c^{trans} - \hat{\mu}_c^{\,trans}}{\sqrt{\hat{\sigma_c^2}^{\,trans} + \epsilon}} + \beta^s$$

$$h_c^{trans} = \frac{h_c^t - \hat{\mu}_c^{\,t}}{\sqrt{\hat{\sigma_c^2}^{\,t} + \epsilon}} \times \sigma_c^{2s} + \mu_c^s$$

Visual Domain Bridge

$$\hat{\mu_c}^{trans} = \alpha \mu_c^{trans} + (1-\alpha)\,\mu_c^s$$
$$\hat{\sigma_c}^{trans} = \alpha \sigma_c^{trans} + (1-\alpha)\,\sigma_c^s$$

Where

$$\mu_c^{trans} = \frac{1}{NHW} \sum\nolimits_{n,h,w} h_{nchw}^{trans}$$

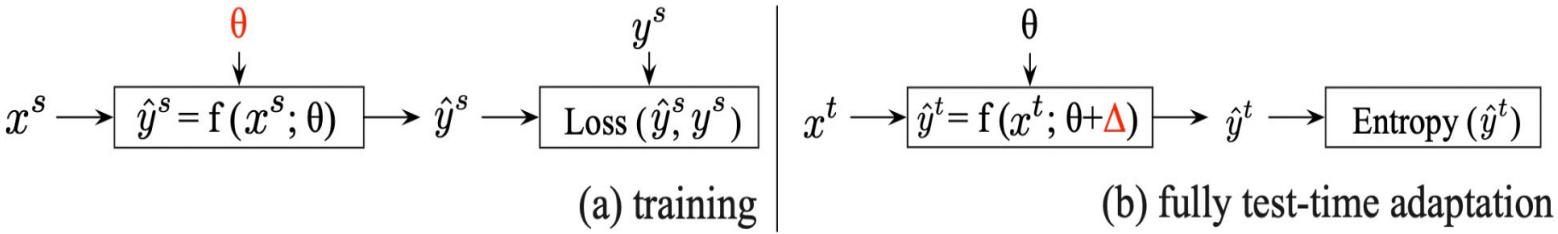
$$\sigma_c^{trans} = \sqrt{\frac{1}{NHW} \sum\nolimits_{n,h,w} (h_{nchw}^{trans} - \mu_c^{trans})^2}$$

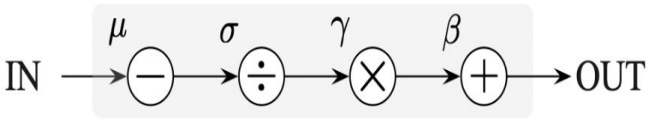
VDB Conclusion

As we can see VDB is transfering the target domain statistic to the source domain, and thus can significantly improve a domain adaptation problem.

TENT

In order to adapt a model to a target distribution, TENT update the affine parameters of the batch norm while minimizing the entropy of the model.





normalization $\mu \leftarrow \mathbb{E}[x_t], \sigma^2 \leftarrow \mathbb{E}[(\mu - x_t)^2]$

transformation $\gamma \leftarrow \gamma + \partial H/\partial\gamma, \beta \leftarrow \beta + \partial H/\partial\beta$

VDB vs TENT ? (personal conclusion)

VDB works well on a few shot problem because it just transfert the statistic of the target domain, at the same time maintain the statistic of the source domain.

For me TENT doesn’t work because it reset the statistic affine of a batch norm and are adapting this statistic with the target domain information, neglecting the source domain statistic.